DISCONTINUITIES OF OPTIMAL CONTROL

by

BALARAM DAS

DEPARTMENT OF MATHEMATICS
INDIAN INSTITUTE OF TECHNOLOGY, KANPUR

MAY, 1983

DISCONTINUITIES OF OPTIMAL CONTROL

A Thesis Submitted
In Partial Fulfilment of the Requirements
for the Degree of
DOCTOR OF PHILOSOPHY

by

BALARAM DAS

to the

DEPARTMENT OF MATHEMATICS

INDIAN INSTITUTE OF TECHNOLOGY, KANPUR

MAY, 1983

freedom from the desire

for an answer is essential

to the understanding of

a problem

– Jiddu Krishnamurty

for

GOURI and BEHARI LAL

to each of whom

I owe more than I know

CERTIFICATE

Certified that the work presented in this thesis, entitled "Discontinuities of Optimal Control", by Sri Balaram Das has been carried out under my supervision and has not been submitted elsewhere for a degree.

Krishna Kumar

Dr. Krishna Kumar
Professor
Department of Aeronautical Engg.
Indian Institute of Technology, Kanpur.

ACKNOWLEDGEMENT

I am sincerely grateful to Prof. Krishna Kumar, a gentleman first and foremost, but for whom this work would have become more difficult. I am indebted to him, in particular, for the unqualified freedom he gave, which made the research for this work proceed in a rather unorthodox manner making it enjoyable all the while and for his patience during the difficult days which saw me through them.

I am also grateful to Prof. P.C. Das who spent some, very cold, winter evenings with me to read the manuscript and give me the benefit of his opinion at several points.

My thanks are also due to Sri Radha Krishna Patnaik, my good friend, who helped me in proof-reading and to Sri G.L. Misra for the nice typing.

10th May 1983 Balaram Das

Table of Symbols

(p.10) denotes page 10 of this thesis.

$\mathbb{R}$	the real line
$\mathbb{R}^n$	n-dimensional Euclidean Space
φ	the null set
∂	boundary
Int	interior
$\bar{A}$	closure of A
$x, x(t)$	the state variable (p.9)
$u, u(t)$	the control (p. 9)
$f(x,u)$	(p. 9)
M	the state space (p.9)
M_x	tangent space to M at $x \varepsilon M$, [3, p.3-18]
M^*_x	cotangent space to M at $x \varepsilon M$, [3, p. 4-3]
TM	tangent bundle of M, [3, p. 3-19]
T^*M	cotangent bundle of M, [3, p. 4-3]
X, X_i	vector fields [3, p. 3-33]
$\Phi(t,\cdot)$, $\Phi_i(t,\cdot)$	one parameter group of diffeomorphisms generated by X, X_i, [3, 5-21]
$[X,Y]$	bracket of X and Y meaning the Lie derivative of Y with respect to X, [3, p. 5-29]
$\lambda, \lambda(t)$	costate variable (p. 24)
π	natural projection associated with the fibre bundle T^*M (p. 25) see also [3, p. 3-11]
$[x,\lambda]$	an element of T^*M (p. 25)
$H([x,\lambda],u)$	the Hamiltonian (p. 25)
Z	a subset of T^*M, (p. 27)

$h([x,\lambda],u)$	$\frac{\partial H}{\partial u}$, (p. 29)
$\varepsilon^{i}_{[x,\lambda]}$	local extremal surface for the point $[x,\lambda]$, (p. 34)
E, E^{a}, E^{b}, E^{i}	extremal surface, (pp. 36, 39, 40, 41)
Ψ	(pp. 36-37)
Σ	a subset of Z, (pp. 40-41)
$L_X(\ )$	Lie derivative of ( ) with respect to the vector field X [3, p. 5-24]
C_i	control region, (pp. 52, 61)
C_{ij}	the set of all points in ∂C_i where the optimal path crosses over from C_i into C_j, (p. 62)
∂^M	(p. 66)
Int^M	(p. 66)
$\pitchfork$	transversal to, (p. 66)
codim	codimension
X'	(p. 101)
Φ'	(p. 101)
Δ_1'	(p. 101)
$\mathrm{Int}(C'_{12})$	(p. 101)
$\mathrm{Int}(C'_{21})$	(p. 101)

CONTENTS

SYNOPSIS

DISCONTINUITIES OF OPTIMAL CONTROL

B. Das
Department of Mathematics,
Indian Institute of Technology, Kanpur.

We consider the time optimal control problem for non-linear systems of the form

$$\frac{dx}{dt} = f(x,u), \qquad (1)$$

here x belongs to an n-dimensional smooth manifold M, f is well defined for our purpose and u is a scalar control with $a \leq u \leq b$ for some real a and b. Our aim is to study the discontinuities of the optimal control. More precisely, we seek answers to the following questions. What are the prin factors that lead to the appearance of discontinuities in the optimal control and which conditions on the system put an upper bound on the number of these discontinuities, regardless of the initial and the target points?

We initiate the study with an investigation to deterr whether the dimensionality of the state space influences th number of discontinuities of the optimal control. We show in systems of the form (1) - where it is possible to have the optimal trajectories as concatenations of the integral curves of a countable family of C^1 vector fields defined or M and where, given $x \in M$, the set of points optimally reachable from x possesses a non-empty interior in M- there

exists an explicit dependence of the number of discontinuiti
of the optimal control on the dimensionality of the state space. This dependence lies in the sense that an upper bound on the number of discontinuities, if it exists, cannot be a number less than $(n-1)$. The analysis also enables us to explain the dependence of the Fel'dbaum's switching theorem, for linear systems, on the dimensions of the state space.

In the second stage we make some geometric constructio with the aim of obtaining a proper representation of the optimal control. For this we utilise the Pontryagin's maximum principle. This requires the definition of a co-sta vector λ and the Hamiltonian

$$H = \langle \lambda, f(x,u) \rangle.$$

The state and co-state variables then satisfy the Hamiltonia equations

$$\dot{x} = \frac{\partial H}{\partial \lambda}$$

$$\dot{\lambda} = - \frac{\partial H}{\partial x}$$

The integral curves of the above system of equations lie in the cotangent bundle $T^{*}M$. Consider all these integral curves along which the Pontryagin's necessary conditions are satisfied. An optimal trajectory in M is the projectio of one such integral curve in $T^{*}M$, we call this integral curve an optimal path. The optimal paths lie in a subset

$Z \subseteq T^*M$ and we take these as the basis of our investigation. It then becomes necessary to represent the optimal control as a function $(x,\lambda) \to u(x,\lambda)$ on Z. This leads us to a partition of Z by the partition sets $\{C_i\}$. C_i, called the control regions, are determined by the fact that at every point in C_i the optimal control is represented by a continuously differentiable function $\mu_i(x,\lambda)$ defined everywhere on C_i. The functions $\{\mu_i\}$ collectively express the optimal control in the form $(x,\lambda) \to u(x,\lambda)$. In the process we also find out the factors which lead to the appearance of discontinuities and non-differentiabilities in the optimal control. We note that this analysis assumes the non-singularity of the system with respect to the maximum principle.

The optimal control suffers a discontinuity or a non-differentiability only when the optimal path crosses over from one control region into another. In the third stage of our analysis we consider the problem of putting an upper bound on the number of such crossings which give rise to non-smoothness in the optimal control. To this end we derive a geometric condition called the transversality condition and show that when a system satisfies the **transversality** condition the total number of discontinuities and non-differentiabilities suffered by the optimal control along an optimal path does not exceed the number $(n-1)$. Finally, we prove that for linear time optimal problems, the transversality condition follows from the assumptions made in the Fel'dbaum's switching theorem.

CHAPTER 0

INTRODUCTION

0-1 Preliminary Remarks

Given a mathematical description of some physical process to be controlled, the objective of the optimal control theory is to determine the control function that steers the state point from a given initial subset to a given target subset of the state space minimizing a pre-assigned cost functional in the process. The control problems that arise while obtaining mathematical models of real world physical problems are mostly non-linear in character. For these problems, an attempt to obtain the optimal control in an analytic form will in general lead us nowhere. The next best thing to do, therefore, is to gather information about the nature of the optimal control function.

If one looks at the solved problems in literature on optimal control theory, the first thing one notices is that, in general, the optimal control suffers discontinuities while steering the state point from the initial position to the final one giving rise to corners in the optimal trajectory in the process. In cases where the optimal control escapes being discontinuous, one can almost always make it so by distancing the boundary points farther apart. This is very interesting (to our way of thinking). Furthermore, it is not

all that easy to gain an intuitive understanding of this tendency to become discontinuous. For, one may be lured into thinking that an endeavour that accomplishes a set goal with a minimum effort has to be a smooth one. The purpose of this work is to glean as much information as we can about the discontinuities of the optimal control.

At the outset we ask the following two questions :

- What makes the optimal control suffer discontinuities?
- How many are they?

The later question as stated above is, however, a bit ambiguous. For, one knows that stretching the boundary points farther apart is very likely to result in more number of discontinuities of the optimal control. A precise form of the later question will be as follows :

Given a set of boundary conditions, how many are the number of discontinuities that the optimal control suffers while steering the state from the initial to the target point?

The question thus formulated embodies some inherent unsurmountable difficulties. For, an attempt to answer it will carry us into the deep and until now unfathomable waters of two point boundary value problem. We will therefore ask a question which is not dependent on the boundary points.

Which conditions, we ask, on the system to be controlled puts an upper bound on the number of discontinuities of the optimal control, no matter how far we travel along any

optimal trajectory from any point in the state space?

In seeking an answer to the above question, we do not propose to exhaust all possibilities. We will not be able to. When the state equations are non-linear, we believe, there are many ways of putting an upper bound on the number of discontinuities of the optimal control. In our attempt to find conditions for an upper bound we will take inspiration from the results due to A.A. Fel'dbaum. Consider the time optimal control problem for linear systems of the form

$$\dot{x} = Ax + bu \qquad -1 \leq u \leq 1 \tag{L}$$

where $x \in \mathbb{R}^n$, A is an nxn real matrix and b is an nx1 real vector. The optimal control has been found to be piecewise constant, it takes on the values +1 and -1 alternately suffering discontinuities in the process. Fel'dbaum [1, p.120; 2, p.143] has shown that in systems where A has all real eigenvalues, the optimal control does not suffer more than (n-1) discontinuities no matter how far the final point lies from the initial point. The above linear problem is a particular case of the more general non-linear problem we propose to investigate and our aim is to explore the possibilities for obtaining similar results for the non-linear problem.

The optimal control theory, as its history shows, has arisen out of problems that are of interest in various

branches of science and engineering. The brachistochrone problem led to the foundation of the theory of calculus of variations. More recently Lawden's work [7] on the optimal navigation of a spececraft from one orbit to another led to the singular optimal control problem. Our own interest in the discontinuities of the optimal control also stems from the Lawden's problem. We want to know whether there exists a bound on the number of impulses used in the optimal orbital transfer (Lawden [7], Marchal [9], Small [10], Edelbaum [20]). This problem, however, will not be taken up in this work.

O-2 Method of Investigation

Before looking at how we propose to get on with the analysis, let us say something about the tools we are going to use. Now, a standard assumption in modern control theory is that the state space is a vector space. This is very natural and valid in many situations. However, as pointed out by Brockett [11], there are very natural control problems in engineering and physics where the state space fails to qualify as a vector space. These problems are therefore not treated by classical control theory. To obviate this difficulty we take the state space in our analysis to be a smooth manifold and use the methods of differential geometry for the investigation. A smooth manifold is only a generalization of $\mathbb{R}^n$ - in the sense that, locally, it looks like

$\mathbb{R}^n$ - in which it is possible to define the process of differentiation. Moreover, this generalization does in no way complicate the nature of the control problem. On the other hand, methods of differential geometry, being compact, give an easy intuitive feeling of the dynamical behaviour of the system under consideration. With regard to the theory of non-linear systems, it is now an established fact that modern differential geometry is uniquely suited to its development [12]. Recent work in geometric control theory by Brockett, Elliott, Krener, Sussmann and others attests to the usefulness of geometric methods [8].

In this work we adhere to the terminology and definitions as given by Michael Spivak in his delightful book on differential geometry [3].

0-3 A View of the Analysis and a Review of the Literature

We propose to consider the time optimal control problem for non-linear systems of the form

$$\frac{dx}{dt} = f(x,u), \tag{0.1}$$

where x belongs to an n-dimensional smooth manifold M, f is well defined for our purpose and u is a scalar control with $a \leq u \leq b$ for some real a and b. The investigation is planned in three stages.

Looking back at Fel'dbaum's switching theorem for linear systems we see that the bound on the number of

discontinuities, when it exists, depends upon the dimensions of the state space. Why does this dimensionality influence the number of discontinuities? In the first stage of our analysis we answer this question. The answer lies in recognizing the fact that when the state space M is paracompact and $n > 1$, the set of points reachable from some $p \varepsilon M$ along the integral curve of a single vector field defined on M does not possess an interior. To endow the reachable set from p with an interior, one has to have more than one vector field defined on M and be able to switch from the integral curve of one vector field to that of another. We show that when the totality of vector fields at our disposal is countable (it is so in Fel'dbaum's case; in fact there are only two vector fields to manoeuvre with) the upper bound, if it exists, on the number of such switchings - required to make the reachable set possess a non-empty interior - is a number not less than (n-1). And this is where the dimensionality of the state space comes into picture.

That the dimensionality of the state space is a governing factor in choosing a family D of vector fields in M, such that the set of points D-reachable from $p \varepsilon M$ (i.e., points in M reachable from p along the integral curve of some vector field in D or along a concatenation of such integral curves; going backwards in time not being allowed) possesses a non-empty interior, has been shown by Sussmann and Jurdjevic [13] in a comprehensive paper on controllability.

In a later paper, Krener [14] generalizes some of these results. In yet another paper Levitt and Sussmann [15] have shown that on every connected paracompact manifold M of class C^k, $k \geq 2$, there exists a set S consisting of two C^{k-1} vector fields on M such that the set of points S-reachable from any $p \in M$ is the entire manifold M and that it is possible to have a bound on the number of switchings. In this context results by Sussmann [16] on local controllability are also interesting.

In the second stage we propose to build a proper **framework** to facilitate the analysis of the discontinuities of the optimal control. To this end we exploit the Pontryagin's maximum principle. Pontryagin's principle converts the optimal control problem into a problem in Hamiltonian control systems by making it necessary to define an n-dimensional co-state vector λ. Now Hamiltonian control systems are constructed on the cotangent bundle of the state space. We therefore take the cotangent bundle of M rather than M itself as our base of operation. This immediately requires the representation of the optimal control as a function on the cotangent bundle in the form $(x,\lambda) \to u(x,\lambda)$. In this context brief comments by Sussmann in one of his papers [17] are quite interesting. In the process of obtaining the optimal control in the form $(x,\lambda) \to u(x,\lambda)$, we also find out the reasons which compel the optimal control to suffer discontinuities and non-differentiabilities.

In the third stage we take up the question of putting an upper bound on the number of discontinuities of the optimal control. As far as we know, except for Fel'dbaum's switching theorem there is little literature on optimal control theory that has direct bearing on this aspect. Hajek in his paper [18] extends the Fel'dbaum's theorem. He considers linear problems of the form (L) with no restrictions on the eigenvalues of A and shows that for every such problem there exists a fundamental length $\varepsilon > 0$ with the property that the optimal control in any time interval of length less than ε suffers at most (n-1) discontinuities. In [19] Sussmann studies systems of the form

$$\dot{x} = f(x) + u\, g(x), \quad -1 \leq u \leq 1.$$

He proves a bang-bang theorem with a priori bounds on the number of switchings.

In our analysis, we actually put a bound on the total number of discontinuities and non-differentiabilities of the optimal control. We show that when System (0.1) satisfies a certain geometric condition called the transversality condition, the total number of discontinuities and non-differentiabilities suffered by the optimal control governing any optimal trajectory in M does not exceed (n-1) no matter how long the time interval is. Finally, we prove that the Linear System (L) satisfies the transversality condition when A has all real eigenvalues.

CHAPTER 1

PREPARATORY ANALYSIS

1-1 Problem Statement

We consider the time optimal control problem for nonlinear systems of the form

$$\frac{d\,x(t)}{dt} = f(x(t),u(t)). \tag{1.1}$$

Here x belongs to an n-dimensional smooth (C^∞) manifold M and the control u belongs to a class U of scalar admissible controls. We define U as follows :

If $\nu \in U$ then

(i) $\nu : I \to \Omega \subseteq \mathbb{R}$ is piecewise continuous where $I \subseteq [0,\infty)$ is an interval and $\Omega \equiv [a,b]$.

(ii) The discontinuities of ν are of the first kind, i.e., we assume the existence of the limits

$$\nu(\tau-0) = \lim_{\substack{t\to\tau \\ t<\tau}} \nu(t) \quad\text{and}\quad \nu(\tau+0) = \lim_{\substack{t\to\tau \\ t>\tau}} \nu(t)$$

at a point of discontinuity τ.

To avoid confusion, we take the value of ν at a point of discontinuity to be its right hand limit, i.e.,

$$\nu(\tau) = \nu(\tau+0).$$

The mapping

$$f : M \times \Omega \to TM$$

(TM is the tangent bundle of M) is twice continuously differentiable with respect to x and u.

Given a control function $t \to u(t)$ defined on an interval I, a trajectory governed by $u(t)$ is a curve $x : I \to M$ that satisfies

$$\dot{x}(t) = f(x(t), u(t))$$

for almost all $t \in I$. Given the System (1.1) with an initial point $x \in M$ and a target point $y \in M$, we say that the control $u : [0,T] \to \Omega$ steers the system from x to y, in time T, if there exists a trajectory $x : [0,T] \to M$ governed by it which satisfies $x(0) = x$ and $x(T) = y$. Furthermore, of all the admissible controls steering the system from x to y, the control u which minimizes the transit time will be called the optimal control and we assume that it is unique. The trajectory governed by an optimal control will be called an optimal trajectory.

As indicated in the introduction, our aim is to study the discontinuities of the optimal control.

1-2 Dependence of the Number of Discontinuities on the Dimensionality of M.

In this section we direct our investigation to study the dependence, if any, of the number of discontinuities of the optimal control on the dimensionality of the state space M. For this purpose we choose a special system from among the systems of the form (1.1). We describe this system below.

First of all we assume M to be paracompact. Let $y \in M$ and let $M(y) \subsetneq M$ be the set of all points in M from which there exist optimal trajectories to y. Next we assume that for every $y \in M$ there exists a function $v_y(x)$,

$$v_y : G(y) \to \mathbb{R},$$

defined on some open subset $G(y) \subsetneq M$ such that $G(y) \supsetneq M(y)$ and v_y when restricted to $M(y)$ acts as the synthesizing function with respect to the target y. In other words the integral curve of the equation

$$\frac{dx}{dt} = f(x, v_y(x)) \qquad (1.2)$$

from any point $x \in (M(y)-\{y\})$ yields the optimal trajectory from x to y. In general $v_y(x)$ is a discontinuous function. We assume that it is differentiable everywhere except, of course, at the points of discontinuity. The number of discontinuities that $v_y(x)$ suffers while steering the system (1.2) from x to y is equal to the number of discontinuities of the optimal control $u(t)$, which governs the optimal trajectory from x to y, in the optimal time interval. We therefore focus our attention on the discontinuities of $v_y(x)$ instead.

It would be useful to interpret the discontinuities of $v_y(x)$ in the following way. As we start from x and proceed along the optimal trajectory, $v_y(x)$ varies smoothly until a point of discontinuity is reached, where it undergoes a jump

and then varies smoothly again. In other words $v_y(x)$ switches from one differentiable function to another suffering a discontinuity in the process and this process possibly gets repeated a number of times before y is reached.

More precisely let D(M) be the set of all differentiable functions defined on M. We assume that there exists a partition of $(M(y)-\{y\})$ by the partition sets $\{S_i\}$, say, such that

(i) S_i are connected and for each S_i it is possible to find a function $v_y^i(x) \in D(M)$, such that

$$v_y^i(x) = v_y(x), \quad \forall\, x \in S_i$$

(ii) and that S_i are maximal with respect to the above property, i.e., there does not exist a connected subset σ_i of $(M(y)-\{y\})$ and a function $h(x) \in D(M)$ with the property that $\sigma_i \supset S_i$ and

$$h(x) = v_y(x) \quad \forall\, x \in \sigma_i.$$

We denote by $D_y(M)$ the collection of all functions $\{v_y^i(x)\}$. (Note : For an S_i it is possible to find more than one function belonging to D(M) which can serve as $v_y^i(x)$. However, in $D_y(M)$ we include one differentiable function for each S_i).

The functions $v_y(x)$ are different for different $y \in M$, so also are the subsets $D_y(M)$. We further assume that there exists a countable subset $D'(M) \subset D(M)$ such that $D_y(M) \subseteq D'(M)$ for all $y \in M$. For each $g(x) \in D'(M)$ the

mapping $f(x,g(x))$ gives rise to a C^1 vector field on M. The collection $D'(M)$ therefore gives rise to a countable family of C^1 vector fields on M. We denote this family by V and by V_y we denote the family of vector fields corresponding to the functions in $D_y(M)$.

The above discussion describes the special case of the System (1.1) we have chosen for analysis in this section. We denote this system by S. S is roughly a subsystem of the family of vector fields V defined on M. With this in mind we define a few terms.

Definition 1

An integral curve of V is a mapping α from a real interval $[t,t']$ into M such that there exists a division $t = t_o < t_1 < \cdots < t_k = t'$ and elements $X_1, X_2, \ldots, X_k$ of V with the property that the restriction of α to $[t_{i-1}, t_i]$ is an integral curve of X_i for each $i = 1,2,\ldots,k$.

Definition 2

Let $x \in M$. A point $y \in M$ is V-reachable from x if there exists $T \geq 0$ and an integral curve $\alpha : [0,T] \to M$ of V such that $\alpha(0) = x$ and $\alpha(T) = y$. Let the set of all points V-reachable from x be denoted by $L_x(V)$.

Definition 3

Let $\alpha : [t_o, t_1] \to M$ be an integral curve of V. We say that $t \in (t_o, t_1)$ is a switching point of α if α is an integral

curve of $X_i \in V$ in $(t-\delta, t]$ and an integral curve of $X_j \in V$ in $[t, t+\varepsilon)$ for some δ, $\varepsilon > 0$ and $X_i \neq X_j$.

Definition 4

Let I^m denote the m-tuple of integers $(i_1, i_2, \ldots, i_m)$. We say that a point $y \in M$ is V-reachable from x with the switching mode I^m if there exist vector fields $X_{i_k} \in V$; $k = 1, 2, \ldots, m$, an integral curve $\alpha : [0,T] \to M$ of V for some $T > 0$ and a division $0 = t_0 < t_1 < \ldots < t_m = T$, such that $\alpha(0) = x$, $\alpha(T) = y$ and α when restricted to the interval (t_{k-1}, t_k) is an integral curve of the vector field X_{i_k}.

Definition 5

Let $x \in M$ and m be an integer. A point $y \in M$ is said to be V-reachable from x with m switchings if there exists an integral curve $\alpha : [0,T] \to M$ of V such that $\alpha(0) = x$, $\alpha(T) = y$ and the interval $(0,T)$ contains m switching points of α.

Given a vector field $X_i \in V$, we denote by $\Phi_i(t, \cdot)$ the one-parameter group of diffeomorphisms generated by X_i [3,p.5-21]. The following theorem is central to the analysis in this chapter.

Theorem 1

Let V be a countable set of C^1 vector fields defined on an n-dimensional manifold M. Let $x \in M$ and $L_x(V)$ have a non-empty interior. If there exists an integer m such that all the points in $L_x(V)$ are V-reachable from x with less than

or equal to m switchings then $m \geq n-1$.

Proof. Suppose that the integer m exists. For an integer i we denote by $L_x(V,i)$ the set of all points in M, V-reachable from x with i switchings. It should be noted that a point $y \in M$ can be reached from x along different integral curves of V with different number of switchings. The sets $L_x(V,i)$, $i = 1,2,\ldots,$ are therefore not necessarily all **disjoint.** Clearly

$$L_x(V) = \bigcup_{i=0}^{m} L_x(V,i).$$

$L_x(V)$ has a non-zero measure [5,p.33]. It therefore follows that there exists an integer k, $0 \leq k \leq m$, for which $L_x(V,k)$ has a non-zero measure. Let I^{k+1} denote the k+1-tuple of integers $(i_1,i_2,\ldots,i_{k+1})$. We denote by $L_x(V,k,I^{k+1})$ the set of all points in M, V-reachable from x with the switching mode I^{k+1} (Def. 4). Clearly

$$L_x(V,k,I^{k+1}) \subseteq L_x(V,k)$$

Moreover the set $L_x(V,k,I^{k+1})$ corresponds to the collection $\{X_{i_1},X_{i_2},\ldots,X_{i_{k+1}}\}$ of (k+1) vector fields belonging to V, in the order given. We note that there is no need for any two of the integers $i_1,i_2,\ldots,i_{k+1}$, except when they are consecutive, to be different. Now V contains a countable number of vector fields. It is therefore possible to have only a countable number of different collections of vector fields where each collection consists of (k+1) vector fields

(not all different) belonging to V. Hence there are only a countable number of sets such as $L_x(V,k,I^{k+1})$ and their union is $L_x(V,k)$ which has a non-zero measure. We therefore conclude that there exists a switching mode J^{k+1} corresponding to the (k+1)-tuple $(j_1,j_2,\ldots,j_{k+1})$ such that the set $L_x(V,k,J^{k+1})$ has a non-zero measure.

$$L_x(V,k,J^{k+1}) \subseteq L_x(V,k) \subseteq L_x(V) \subseteq M$$

The points of $L_x(V,k,J^{k+1})$ are of the form

$$\Phi_{j_{k+1}}(t_{k+1},\Phi_{j_k}(t_k,\ldots,\Phi_{j_1}(t_1,x)\ldots))$$

for some real $t_1,t_2,\ldots,t_{k+1}$, all positive. Let the subset $A \subseteq \mathbb{R}^{k+1}$ be such that

(i) $(t_1,t_2,\ldots,t_{k+1}) \in A \Longrightarrow t_1 > 0,\ t_2 > 0,\ldots,t_{k+1} > 0$

(ii) A consists of all points in $\mathbb{R}^{k+1}$ at which the function F given by

$$F(t_1,t_2,\ldots,t_{k+1})$$

$$= \Phi_{j_{k+1}}(t_{k+1},\Phi_{j_k}(t_k,\ldots,\Phi_{j_1}(t_1,x)\ldots))$$

is defined.

Clearly A is an open subset of $\mathbb{R}^{k+1}$ and

$$F(A) = L_x(V,k,J^{k+1})$$

Since the vector fields $X_{j_1},X_{j_2},\ldots,X_{j_{k+1}}$ are all C^1-differentiable, the mapping $F : A \to M$ is also C^1-differentiable

F maps a submanifold of $\mathbb{R}^{k+1}$ onto a subset of M having a non-zero measure. We therefore conclude that

$$k+1 \geq n \Longrightarrow k \geq n-1$$

$$\Longrightarrow m \geq n-1.$$

(Let Xand Y be differentiable manifolds of dimensions n and m respectively with $n < m$. Let $f : X \to Y$ be C^1-differentiable, then f(X) has measure zero in Y. See [5,p.31] for a proof.) This completes our proof.

Now a few remarks concerning this theorem :

Remark 1. Let D denote a set of complete analytic vector fields defined on M. A vector field X is complete if the integral curves of X are defined for all real t. Let T(D) be the smallest Lie algebra of analytic vector fields on M which contains D (the bracket operation being the Lie product of vector fields). Sussmann and Jurdjevic [13] have shown that a necessary and sufficient condition for $L_x(D)$ to have a non-empty interior in M is that dim T(D)(x) = dim M, where T(D)(x) is the subspace of M_x spanned by the vectors X(x), $X \varepsilon T(D)$.

Remark 2. It is clear that the above theorem revolves around the fact that $L_x(V)$ has a non-empty interior. By assuming this we have ensured that the reachable set has an n-dimensional freedom. That this dimensional extension gives rise to the necessity of at least (n-1) switchings, can be

realized intuitively also. For, were we to allow no switchings, the reachable set from the point x will clearly be one of a single dimension (excluding the point x), being the countable union of smooth integral curves of different members of V through the point x; and if one switching were allowed, the reachable set would at best be of two dimensions. Switchings are therefore necessary to increase the dimensions of the reachable set.

The optimal trajectories of the system S are nothing but integral curves of V. While proving the theorem, above, we had the freedom to switch from any member of V to any other member at any instant. Let this type of switchings be called "free switchings".

Consider the optimal problem for the system S, in an attempt to reach a point y from x ($x,y \in M$) optimally, we can not afford free switchings, for, we have to satisfy the constraints of optimality. Let $\alpha : [0,T] \to M$ be the optimal trajectory from x to y for the system S. α is an integral curve of V, in particular of V_y and $[0,T]$ is the optimal time interval. Let the function $u(t) : [0,T] \to \mathbb{R}$ be the governing optimal control. Now it may be possible to construct α in different ways using different subsets of V. However, considering the way V_y is constructed there exists a way of constructing α using vector fields belonging to V_y such that every switching point $\tau \in (0,T)$ of α is a point of discontinuity of $u(t)$. We consider all optimal trajectories of the

system S to have been constructed in this way and by "optimal switchings" we mean the switchings encountered along an optimal trajectory. Optimal switchings are therefore a particular case of free switchings. We have therefore proved the following corollary.

Corollary

Given any $x \in M$ let the set of points reachable from x along optimal trajectories of S have a non-empty interior. Let N be the set of integers such that $k \in N \Longrightarrow$ there exists an optimal trajectory of S along which the optimal control suffers k discontinuities and we take into account all possible optimal trajectories of the system S while constructing N. Then the least upperbound of the set N, if it exists, is a number greater than or equal to (n-1) where n is the number of dimensions of the state space M.

The above analysis also explains the reason why the optimal control for the system S suffers discontinuity. Discontinuities are necessary to increase the dimensionality of the optimally reachable set.

To illustrate the significance of the corollary let us consider the linear time optimal problems of the form

$$\dot{x} = Ax + bu, \qquad -1 \leq u \leq 1 \qquad \text{(L)}$$

Here $x \in \mathbb{R}^n$, A is an $n \times n$ real matrix and b an $n \times 1$ real vector. We assume that (L) is controllable. An assertion about the

number of discontinuities of the optimal control is spelled out in Fel'dbaum's switching theorem [1,p.120]. We state it as follows :

Let all the eigenvalues of A be real. Then, the optimal control is piecewise constant, takes on only the values +1 and -1 and does not have more than (n-1) discontinuities.

If, on the other hand, A has complex eigenvalues it is not possible to put a bound on the number of discontinuities [18].

The optimal trajectories of (L) are nothing but integral curves of $V = \{Z_1, Z_2\}$. Z_1 and Z_2 are C^∞ vector fields defined on $\mathbb{R}^n$ by the rule

$$Z_1(x) = Ax + b \text{ and } Z_2(x) = Ax-b.$$

Controllability of L ensures that given $x \in \mathbb{R}^n$, the set of points reachable from x along optimal trajectories of L has a non-empty interior - in fact the set is equal to $\mathbb{R}^n$. The above corollary now applicable to (L) explains the dependence of Fel'dbaum's theorem on the dimensions of the state space, in particular the dependence on the number (n-1). A number less than (n-1) can never be an upper bound and it so happens that, when the eigenvalues of A are all real a point $y \in L_x(V)$ is V-reachable from x with no more than (n-1) switchings [18].

1-3 A Theorem on Free Switchings.

In the wake of Theorem 1 the following question arises naturally. Is it possible to find a countable family V of

vector fields on M such that all the points in $L_x(V)$ are V-reachable from x with no more than (n-1) switchings? In the discussion for linear systems we saw that if $V = \{Z_1, Z_2\}$ and all the eigenvalues of A are real, then V possesses this property. We also believe, there are many ways of choosing a countable collection, possessing the above property, from the set of all non-linear vector fields defined on M. One such collection for example, is the set V whose elements give rise to a co-ordinate system on M. We conclude this chapter with a theorem which proves this intuitively obvious fact.

Theorem 2

Let M be an n-dimensional analytic manifold. Let V be the set of n vector fields $\{X_1, X_2, \ldots, X_n\}$ defined on M which satisfy, the following conditions :

(i) X_i are analytic and complete, $i = 1, 2, \ldots, n$.

(ii) X_i are linearly independent everywhere on M.

(iii) $[X_i, X_j] = 0$ on M for $1 \leq i, j \leq n$.

Then a point $y \in L_x(V)$ is V-reachable from x with no more than (n-1) switchings.

Proof. We note that $L_x(V)$ has a non-empty interior for all $x \in M$ [13]. Condition (iii) further ensures the commutativity of the vector fields in V. If X_i and X_j belong to V and generate the one-parameter groups of diffeomorphisms $\Phi_i(t,\cdot)$ and $\Phi_j(t,\cdot)$ respectively then for any $x \in M$

$$\Phi_i(t_i,\Phi_j(t_j,x)) = \Phi_j(t_j,\Phi_i(t_i,x)) \qquad (*)$$

Let $y \varepsilon L_x(V)$. Then there is an integral curve $t \to \alpha_1(t)$ of V which connects x to y. Let this integral curve be of the form

$$x = \alpha_1(0)$$

$$y = \alpha_1(t_1+t_2+\dots+t_k)$$

$$= \Phi_{i_k}(t_k,\Phi_{i_{k-1}}(t_{k-1},\dots,\Phi_{i_1}(t_1,x)\dots))$$

corresponding to the vector fields $X_{i_1}, X_{i_2}, \dots, X_{i_k}$ in V. If $X_{i_1} \equiv X_{i_m}$, $1 < m \leq k$, then from Equation (*) it follows that there exists an integral curve α_2 of V, connecting x to y, which has the form

$$x = \alpha_2(0)$$

$$y = \alpha_2(t_1+t_2+\dots+t_k)$$

$$= \Phi_{i_k}(t_k,\dots,\Phi_{i_{m+1}}(t_{m+1},\Phi_{i_{m-1}}(t_{m-1},\dots,\Phi_{i_1}(t_1+t_m,x)$$

$$\dots))\dots).$$

Applying the above technique repeatedly we can therefore find an integral curve α of V connecting x to y such that

(a) $\alpha(0) = x$, $\alpha(t_1+t_2+\dots+t_k) = y$ and

(b) given $X_i \varepsilon V$, α is an integral curve of X_i in a connected subinterval of $[0,t_1+t_2+\dots+t_k]$.

However as there are only n vector fields in V, it follows that α can contain at most (n-1) switchings. This concludes the proof and the chapter.

CHAPTER 2

CONTROL REGIONS

2-1 Preliminaries

In this chapter we make some geometric constructions with a view to facilitate the interpretation of discontinuities of the optimal control. We consider the time optimal control problem for systems

$$\frac{d\,x(t)}{dt} = f(x(t),\, u(t)) \tag{2.1}$$

as defined in Section (1-1).

As $f(x(t),u(t))$ is nonlinear, finding the optimal control $t \to u(t)$ and the corresponding optimal trajectory for the System (2.1) tantamounts to solving a two point nonlinear boundary value problem. We will not attempt anything of this kind. Given the initial and the target points in the state space M we assume that a unique optimal control exists for the system and therefore satisfies the necessary conditions of Pontryagin's maximum principle. We exploit this principle to our advantage.

The maximum principle first of all requires defining a costate variable as follows :

Let (x,V) be a co-ordinate system around a point $p \in M$ Consider the mapping $f : M \times \Omega \to TM$. Let $\alpha \in \Omega$, then the restriction of f to $M \times \{\alpha\}$

$$f : M \times \{\alpha\} \to TM$$

is a vector field on M. Relative to the coordinates $x_1, x_2, \ldots, x_n$ we can write

$$f(\xi,\alpha) = \sum_{i=1}^{n} f_i(\xi,\alpha) \frac{\partial}{\partial x_i} \quad \forall \quad \xi \in V.$$

Let u(t) be an admissible control defined on an interval I. Let $x : I \to M$ be a trajectory governed by u(t). Then the costate variable for the pair (x(t),u(t)) is a map $t \to \lambda(t)$ defined for all $t \in I$ such that $\lambda(t) \in M^*_{x(t)}$, for all $t \in I$, and satisfies the costate equation

$$\frac{d\lambda_i}{dt} = - \sum_{j=1}^{n} \lambda_j \frac{\partial f_j(x,u)}{\partial x_i} \tag{2.2}$$

where $\lambda = \sum_{i=1}^{n} \lambda_i \, dx_i$.

The system of Equations (2.2) are linear and homogeneous and therefore have a unique solution through any initial point.

Definition 1

Given an admissible control $u = u(t)$ defined on an interval I, a curve $(x,\lambda) : I \to T^*M$ that satisfies Equations (2.1) and (2.2) (with the given control) for almost all $t \in I$ will be called an <u>integral path governed by u(t)</u>.

It is clear that trajectories and integral paths governed by admissible controls are continuous, and

differentiable almost everywhere, in the domain of definition. The trajectories in M are natural projections of integral paths in T^*M. By natural projection we always mean the projection

$$\pi : T^*M \to M$$

associated with the fibre bundle T^*M. Let us represent an element of T^*M by the symbol $[x,\lambda]$ then $\pi\,[x,\lambda] = x \in M$ and λ is the corresponding cotangent vector in M_x^*. (Note : A closed interval in the real line is also represented by square brackets e.g. the interval $[a,b]$. This, however, will not create any confusion as the context will clarify the symbol.)

The Hamiltonian for the System (2.1) is a function

$$H : T^*M \times \Omega \to \mathbb{R}$$

defined by the relation

$$H([x,\lambda],u) = \langle \lambda, f(x,u)\rangle. \qquad (2.3)$$

The Pontryagin's maximum principle says: in order that an admissible control u(t) and a trajectory x(t) governed by it be time optimal, it is necessary that there exists a non-trivial costate variable $\lambda(t)$ corresponding to the pair (x(t),u(t)) such that

$$H([x(t),\lambda(t)],u(t)) = \max_{v \in \Omega} H([x(t),\lambda(t)],v) \qquad (2.4)$$

and

$$H([x(t),\lambda(t)]\ ,\ u(t)) \geq 0 \tag{2.4$'$}$$

for almost all t in the domain of interest.

Definition 2

An integral path $t \to [x(t),\lambda(t)] \in T^*M$ governed by the control $u(t)$ will be called an extremal integral path if $\lambda(t) \neq 0\ \forall\ t \in I$, and if for the pair $([x(t),\lambda(t)]\ ,u(t))$ Relations (2.4) and (2.4)$'$ are satisfied almost everywhere on I, where I is the domain of definition of the integral path.

It now follows from the maximum principle that an optimal trajectory in M is the projection under π of some extremal integral path in T^*M. We therefore make the following definition.

Definition 3

An extremal integral path in T^*M whose projection under π is an optimal trajectory in M will be called an optimal path.

Clearly the control $u(t)$ which governs an optimal path is an optimal control.

Now consider an optimal path defined on an interval I. Let $\tau \in I$ be the instant when the optimal path passes through the point $[X,\lambda] \in T^*M$. Let $\nu \in \Omega$ be the value of the governing optimal control at $t = \tau$. Then the following equality is satisfied

$$H([x,\lambda],\nu) = \max_{v \in \Omega} H([x,\lambda],v) \tag{2.5}$$

This relation has nothing to do with time. What we want to stress is that whenever an optimal path passes through a point $[X,\lambda] \in T^*M$, the value of the optimal control there depends only on the point in question. The value of time when we reach the point in question is immaterial. For, we may start from different points lying on the same optimal path, corresponding to different initial points on the same optimal trajectory, and reach $[x,\lambda]$ at different epochs, in all cases, however, we reach the point $[x,\lambda]$ with the same value of the optimal control. If therefore we were to take the optimal path rather than the optimal trajectory as our basis of investigation, it would be wiser to seek the optimal control in the form

$$u : T^*M \to \mathbb{R}. \tag{2.6}$$

We encounter a few difficulties here. It may not be possible to have an optimal path through every point in T^*M. For example, there may exist points $[X,\lambda] \in T^*M$ such that $H([x,\lambda],v) < 0$ for all $v \in [a,b]$. Owing to the relation (2.4)', no optimal paths can pass through such points. Let $Z \subseteq T^*M$ be the maximal subset such that it is possible to have an optimal path through every point in Z. The optimal control should therefore be sought in the form

$$u : Z \to \mathbb{R} \tag{2.6'}$$

Finally, let us make ourselves clear with regard to an important limitation before proceeding further. Our aim, as we have said before, is to study the discontinuities of the optimal control to which end we are interested in obtaining the optimal control in the form (2.6)'. We plan to do this with the help of the maximum principle. Now the maximum principle, being a first order necessary condition, will not always be able to specify the optimal control uniquely, even if we were able to ascertain its existence a priori. At times it may be possible to get over this difficulty with the help of higher order necessary conditions. We however, will not get into these complications as they have a distinctly different flavour from what we aim at. We call an optimal control problem "singular with respect to the maximum principle" or in short "singular" if there arise situations where it is not possible to obtain definite information about the optimal control (when it exists), in terms of state and co-state variables, through the maximum principle. Henceforth our analysis will be confined to non-singular problems only. Rules out the interesting cases

2-2 Extremal Surfaces

The Hamiltonian is twice continuously differentiable with respect to the parameter $u \in \Omega \equiv [a,b]$. Therefore the partial derivative $\dfrac{\partial H([x,\lambda],u)}{\partial u}$ is defined for all $[x,\lambda] \in T^*M$ and for all $u \in [a,b]$. At $u = a$(or b), the

smoothness of H means that it can be extended to a smooth function $\bar{H}$ defined in an open neighbourhood of a (or b) ε $\mathbb{R}$; we define $\left.\frac{\partial H([x,\lambda],u)}{\partial u}\right|_{u=a(\text{or } b)}$ to be the derivative $\left.\frac{\partial \bar{H}([x,\lambda],u)}{\partial u}\right|_{u=a(\text{or } b)}$. It can be shown that this definition yields a unique value of the derivative at u = a (or b). For the sake of convenience we write

$$\frac{\partial H([x,\lambda],u)}{\partial u} \equiv h\,([x,\lambda],u) : T^*M \times [a,b] \to \mathbb{R}. \qquad (2.7)$$

For a while let us suppose that the global maximum of $H([x,\lambda],u)$ considered as a function of u in $[a,b]$ is attained in the interior (a,b). If then we succeed in obtaining the optimal control in the form $[x,\lambda] \to u([x,\lambda])$, the equation

$$h([x,\lambda],\ u([x,\lambda])) = 0$$

should necessarily be satisfied for all $[x,\lambda] \varepsilon Z$. One way therefore, to find the optimal control in the form (2.6)' is to seek an explicit solution in the form $u = u([x,\lambda])$ for the equation

$$h([x,\lambda],u) = 0. \qquad (2.8)$$

Let $[x_o,\lambda_o]$ be a point in the cotangent bundle of M. Suppose the Hamiltonian $H([x_o,\lambda_o],u)$ considered as a function of u has only non-degenerate critical points in the interval $[a,b]$. Since the non-degenerate critical points are isolated, only a finite number of such points can occur in the compact

interval $[a,b]$. Let the points $u_{01}, u_{02},\ldots,u_{0\ell}$, in the open interval (a,b) be the non-degenerate critical points of $H([x_o,\lambda_o],u)$. Then at each point $([x_o,\lambda_o],u_{0i}) \varepsilon\ T^*M \times (a,b)$; $i = 1,2,\ldots,\ell$, we have

(i) $h([x_o,\lambda_o],u_{0i}) = 0 \qquad i = 1,2,\ldots,\ell$;

(ii) $h_u([x_o,\lambda_o], u_{0i}) \neq 0 \qquad i = 1,2,\ldots,\ell$.

Moreover $h([x,\lambda],u)$ possesses continuous partial derivatives with respect to all its arguments. It then immediately follows from the implicit function theorem that for each u_{0i}, $i = 1,2,\ldots,\ell$, there exists an open set σ_i in T^*M containing the point $[x_o,\lambda_o]$ such that the equation $h([x,\lambda],u) = 0$ has a unique continuously differentiable solution $u = u_i([x,\lambda])$ defined on σ_i which satisfies $u_{0i} = u_i([x_o,\lambda_o])$.

For a possible optimal path through $[x_o,\lambda_o]$, the value of the optimal control there is given by the value at $[x_o,\lambda_o]$ of one of the functions $\{u_1([x,\lambda]),\ldots,u_\ell([x,\lambda])\}$. The appropriate function u_j is determined by the relation

$$H([x_o,\lambda_o], u_j([x_o,\lambda_o])) > H([x_o,\lambda_o], u_i([x_o,\lambda_o])) \quad (2.9)$$

$$i = 1,2,\ldots,\ell;\ i \neq j$$

The Relation (2.9) has to undergo a slight modification if now we reject the assumption that the global maximum of $H([x,\lambda],u)$ considered as a function of u in $[a,b]$ is attained in the interior (a,b). To see this, define two constant

functions $u_{\ell+1}$ and $u_{\ell+2}$ on T^*M by the relations

$$u_{\ell+1}([x,\lambda]) = a \,;\; u_{\ell+2}([x,\lambda]) = b \qquad (2.10)$$

The value of the optimal control at $[x_o,\lambda_o]$ will then be given by the function $u_k([x,\lambda])$, for some integer k in $1 \le k \le \ell+2$, determined by the relation

$$H([x_o,\lambda_o], u_k([x_o,\lambda_o])) > H([x_o,\lambda_o], u_i([x_o,\lambda_o]));\quad (2.11)$$

$$i = 1,2,\ldots,\ell+2,\; i \neq k.$$

We say that the function u_k defined on the open set σ_k represents the optimal control at $[x_o,\lambda_o]$.

A tacit assumption that $H([x_o,\lambda_o], u)$ considered as a function of u has a unique global maximum in $[a,b]$ has been made while writing the Relation (2.11). At a later stage we analyse the case when the above function has multiple global maxima in $[a,b]$.

We say that the function $\nu : S \to \mathbb{R}$, $S \subseteq T^*M$, represents the optimal control in a subset $\sigma \subseteq S$ if for every $[x,\lambda] \in \sigma$

$$a \le \nu([x,\lambda]) \le b$$

and

$$H([x,\lambda], \nu([x,\lambda])) = \max_{v\in[a,b]} H([x,\lambda],v)$$

Relation (2.11) however implies a stronger result if we assume that either the points a,b are not critical points

of the function $H([x_o,\lambda_o], u)$ or if they are critical points they are non-degenerate and the corresponding solutions (through implicit function theorem) to the equation $h([x,\lambda],u) = 0$ are given by $u = u_{\ell+1}([x,\lambda])$ and $u = u_{\ell+2}([x,\lambda])$. We proceed to obtain this result in the Lemma below

Lemma 1

If the function $u_k([x,\lambda])$ represents the optimal control at $[x_o,\lambda_o] \varepsilon T^*M$, then there exists a neighbourhood N of $[x_o,\lambda_o]$ such that $u_k([x,\lambda])$ represents the optimal control at every point in N.

Proof. We must show that if $u_k([x,\lambda])$ satisfies the Relation (2.11) then there exists a neighbourhood N of $[x_o,\lambda_o]$ such that for all $[x,\lambda] \varepsilon N$

$$\left.\begin{array}{ll} \text{(i)} & a \leq u_k([x,\lambda]) \leq b \\ \text{(ii)} & H([x,\lambda],u_k([x,\lambda])) = \max\limits_{v\varepsilon[a,b]} H([x,\lambda],v) \end{array}\right] \quad \text{(A)}$$

Let $H([x_o,\lambda_o], u_k([x_o,\lambda_o]))-H([x_o,\lambda_o],u_i([x_o,\lambda_o])) = \delta_i ; i \neq k.$ Then $\delta_i > 0$ as is evident from Relation (2.11). Let η be a neighbourhood of $[x_o,\lambda_o]$ such that it is contained in all σ_i. Let $N_i \subseteq \eta$, $i = 1,2,\ldots,\ell+2$, be neighbourhoods of $[x_o,\lambda_o]$ such that for all $[x,\lambda] \varepsilon N_i$, $i = 1,2,\ldots,\ell+2$, we have

(i) $a \leq u_k([x,\lambda]) \leq b$

(ii) $|H([x_o,\lambda_o],u_k([x_o,\lambda_o]))-H([x,\lambda],u_k([x,\lambda]))| < \frac{\delta_i}{3}$

and

$$|H([x_o,\lambda_o],u_i([x_o,\lambda_o]))-H([x,\lambda],\ u_i([x,\lambda]))| < \frac{\delta_i}{3}\ .$$

These are possible because u_k and H are continuous functions on T^*M. For all $[x,\lambda] \in N_i$ we therefore have

$$\begin{aligned}
&H([x,\lambda],u_k([x,\lambda])) - H([x,\lambda],\ u_i([x,\lambda])) \\
&\geq H([x_o,\lambda_o],\ u_k([x_o,\lambda_o]))-H([x_o,\lambda_o],u_i([x_o,\lambda_o])) \\
&\quad -|H([x_o,\lambda_o],u_k([x_o,\lambda_o]))-H([x,\lambda],u_k([x,\lambda]))| \\
&\quad - |H([x,\lambda],\ u_i([x,\lambda])) - H([x_o,\lambda_o]\ ,\ u_i([x_o,\lambda_o]))| \\
&\geq \delta_i - \frac{\delta_i}{3} - \frac{\delta_i}{3} > 0
\end{aligned}$$

or $H([x,\lambda],u_k([x,\lambda])) > H([x,\lambda],u_i([x,\lambda]))\ \forall\ [x,\lambda] \in N_i$.

Let $N = N_1 \cap N_2 \cap \cdots \cap N_{k-1} \cap N_{k+1} \cap \cdots \cap N_{\ell+2}$. Then for all $[x,\lambda] \in N$ we have

$$\left.\begin{aligned}
&\text{(i)} \quad a \leq u_k([x,\lambda]) \leq b \\
&\text{(ii)} \quad H([x,\lambda],u_k([x,\lambda])) > H([x,\lambda],\ u_i([x,\lambda]))
\end{aligned}\right] \quad \text{(B)}$$

$$i = 1,2,\ldots,\ell+2;\ i \neq k.$$

Now the functions $u_i([x,\lambda])$, $i = 1,2,\ldots,\ell$, satisfy the equation $h([x,\lambda],u) = 0$ at all points $[x,\lambda] \in N$. Therefore for every $[x,\lambda] \in N$ the values of u at which $H([x,\lambda],u)$ considered as a function of u in $[a,b]$ attains its extreme values are determined by the functions $\{u_i\}$. That all the extreme values are accounted for can be ensured by choosing N sufficiently small. Therefore Relations (B) imply Relations (A).

We now make some geometric constructions. Take an arbitrary point $[x',\lambda'] \in T^*M$. Let the points $u_1', u_2', \ldots,$ in the closed interval $[a,b]$ be the non-degenerate critical points of $H([x',\lambda'],u)$ considered as a function of u. Then as we have seen before, for each u_i', $i = 1,2,\ldots,$ there exists an open set σ_i' in T^*M containing the point $[x',\lambda']$ such that the equation $h([x,\lambda],u) = 0$ has a unique continuously differentiable solution $u = u_i([x,\lambda])$ defined on σ_i' which satisfies $u_i' = u_i([x',\lambda'])$. We define subsets $\varepsilon^i_{[x',\lambda']} \subseteq T^*M \times \mathbb{R}$ in the following way.

$$\varepsilon^i_{[x',\lambda']} = \{([x,\lambda],u) \in T^*M \times \mathbb{R} : [x,\lambda] \in \sigma_i' \text{ and }$$

$$u = u_i([x,\lambda])\}; \quad i = 1,2,\ldots \qquad (2.12)$$

Then $\varepsilon^i_{[x',\lambda']}$ are $2n$-dimensional submanifolds of $T^*M \times \mathbb{R}$. We call these submanifolds "local extremal surfaces for the point $[x',\lambda']$". Although for each u_i' it is possible to choose different σ_i' and therefore construct different $\varepsilon^i_{[x',\lambda']}$, we choose just one σ_i' and construct just one $\varepsilon^i_{[x',\lambda']}$ for each u_i'. We extend this process and construct local extremal surfaces for all points $[x,\lambda] \in T^*M$.

Let $\varepsilon^i_{[x_1,\lambda_1]}$ and $\varepsilon^j_{[x_2,\lambda_2]}$ be two extremal surfaces for the points $[x_1,\lambda_1]$ and $[x_2,\lambda_2]$ respectively.

$$\varepsilon^{i}_{[x_1,\lambda_1]} = \{([x,\lambda],u) \in T^{*}M \times \mathbb{R} : [x,\lambda] \in \sigma_{1i} \text{ and } , u = u_{1i}([x,\lambda])\},$$

$$\varepsilon^{j}_{[x_2,\lambda_2]} = \{([x,\lambda],u) \in T^{*}M \times \mathbb{R} : [x,\lambda] \in \sigma_{2j} \quad u = u_{2j}([x,\lambda])\}.$$

Here the open sets σ_{1i}, σ_{2j} and the continuously differentiable functions u_{1i} and u_{2j} have meaning as in (2.12). We say that $\varepsilon^{i}_{[x_1,\lambda_1]}$ overlaps $\varepsilon^{j}_{[x_2,\lambda_2]}$ (or equivalently $\varepsilon^{j}_{[x_2,\lambda_2]}$ overlaps $\varepsilon^{i}_{[x_1,\lambda_1]}$) if

(i) $\sigma_{1i} \cap \sigma_{2j} \neq \varphi$ and

(ii) there exists a continuously differentiable function $u_{1i2j} : \sigma_{1i} \cup \sigma_{2j} \to \mathbb{R}$ given by

$$u_{1i2j} = \begin{cases} u_{1i}([x,\lambda]) \ \forall \ [x,\lambda] \in \sigma_{1i} \\ u_{2j}([x,\lambda]) \ \forall \ [x,\lambda] \in \sigma_{2j}. \end{cases}$$

Moreover the surface

$$\varepsilon^{ij}_{[x_1,\lambda_1],[x_2,\lambda_2]} = \{([x,\lambda],u) \in T^{*}M \times \mathbb{R} : [x,\lambda] \in \sigma_{1i} \cup \sigma_{2j} \text{ and } u = u_{1i2j}([x,\lambda])\}$$

is said to be the join of $\varepsilon^{i}_{[x_1,\lambda_1]}$ and $\varepsilon^{j}_{[x_2,\lambda_2]}$.

The local extremal surfaces have the following properties :

(i) Two different local extremal surfaces for the same point do not intersect. Suppose contrary, $\varepsilon^1_{[x',\lambda']} \cap \varepsilon^2_{[x',\lambda']} \neq \varphi$, and let $([x,\lambda],u)$ belong to this intersection. Then through the point $([x,\lambda],u)$, the equation $h([x,\lambda],u) = 0$ has two solutions viz. $u = u_1([x,\lambda])$ and $u = u_2([x,\lambda])$. This violates the conclusions of the implicit function theorem.

(ii) Let $\varepsilon^i_{[x_1,\lambda_1]}$ and $\varepsilon^j_{[x_2,\lambda_2]}$ be two local extremal surfaces for the points $[x_1,\lambda_1]$ and $[x_2,\lambda_2]$ respectively. Then $\varepsilon^i_{[x_1,\lambda_1]}$ and $\varepsilon^j_{[x_2,\lambda_2]}$ are either disjoint or overlapping. For if the two surfaces have common points and fail to overlap, the conclusions of implicit function theorem are again violated.

So far we have been carrying out local analysis and what is important is that it could have been carried out locally only. But local extremal surfaces are submanifolds of the same dimension, 2n, everywhere. There is therefore no difficulty in joining the overlapping surfaces together. If now we join all overlapping local extremal surfaces together, we would end up with a number of 2n-dimensional submanifolds in $T^*M \times \mathbb{R}$. We call these submanifolds "extremal surfaces". We stipulate each extremal surface to be maximal with respect to connectedness.

Let Ψ denote the projection

$$\Psi : T^*M \times \mathbb{R} \to T^*M$$

defined by

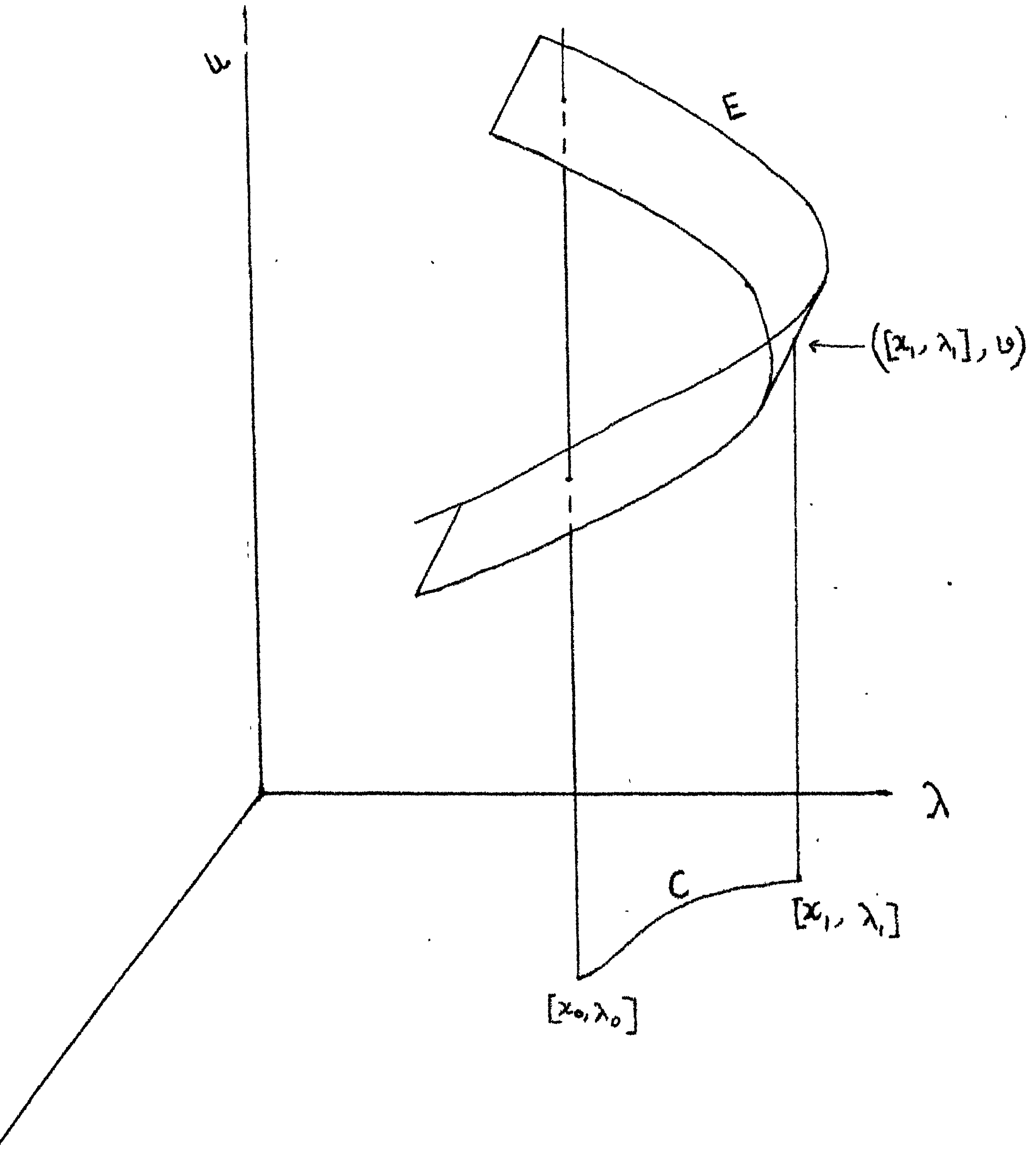
$$\Psi([x,\lambda],u) = [x,\lambda].$$

The extremal surfaces as constructed above have the following two properties :

(i) No two extremal surfaces intersect. This follows directly from the properties of local extremal surfaces.

(ii) Let E be an extremal surface. Let $[x,\lambda] \in T^*M$. Then there does not exist more than one real u such that $([x,\lambda],u) \in E$.

Proof. Suppose the aforesaid property is false. Let E bend over itself once. In other words we can find points of the type $[x_o, \lambda_o] \in T^*M$ such that there exist exactly two real numbers, say, u_1 and u_2 with the property that $([x_o,\lambda_o],u_i) \in E$ for $i = 1,2$ and that we cannot find more than two such real numbers for any point belonging to T^*M. It therefore follows that through the point $[x_o,\lambda_o] \in T^*M$ there exists a curve lying in T^*M such that as we proceed along this curve from the point $[x_o,\lambda_o]$, corresponding to every point on the curve there exist two points on the extremal surface E until we reach a point $[x_1,\lambda_1]$ corresponding to which there exists just one point, say, $([x_1,\lambda_1],v) \in E$. That such curve exists can be ensured by the connectedness of E. Denote by C the set of all points on this curve in between $[x_o,\lambda_o]$ and $[x_1,\lambda_1]$. Let N be a neighbourhood of $([x_1,\lambda_1],v)$ in $T^*M \times \mathbb{R}$. We can choose N such that $\varepsilon = N \cap$ is a local extremal surface for the point $[x_1,\lambda_1]$ -through


u
E
$([x_1, \lambda_1], v)$
λ
C
$[x_1, \lambda_1]$
$[x_0, \lambda_0]$

$([x_1,\lambda_1],v)$. This is possible because extremal surfaces are constructed by joining overlapping local extremal surfaces. Let $N' = \Psi(N)$ and $N'' = (N' \cap C) - \{[x_1,\lambda_1]\}$. Then for every point $[x,\lambda] \in N''$ we can find two real numbers v_1 and v_2, say, such that $([x,\lambda],v_i)$ belongs to ε, $i = 1,2$. This, however, contradicts the fact that ε is a local extremal surface. E, therefore, cannot bend over itself, or for that matter it cannot bend over itself more than once. Property (ii) is therefore true.*

Consider an extremal surface E. The set $\Psi(E) = S$ is open in T^*M. It follows from our previous discussion that there exists a continuously differentiable function $\nu([x,\lambda])$, defined on S satisfying $h([x,\lambda], \nu([x,\lambda])) = 0$ everywhere on S, so that E can be written in the form

$$E = \{([x,\lambda],u) \in T^*M \times \mathbb{R} : [x,\lambda] \in S \text{ and } u = \nu([x,\lambda])\}. \tag{2.13}$$

That $\nu([x,\lambda])$ is well defined follows from the Property (ii) of the extremal surfaces and that it is continuously differentiable and satisfies $h([x,\lambda], \nu([x,\lambda])) = 0$ on S follows from the fact that extremal surfaces are locally equivalent to local extremal surfaces.

Define two more hypersurfaces in $T^*M \times \mathbb{R}$

$$E^a = \{([x,\lambda],u) \in T^*M \times \mathbb{R} : [x,\lambda] \in T^*M \text{ and } u = \nu_a([x,\lambda]) = a\} \tag{2.13a}$$

* See Appendix -A

$$E^b = \{([x,\lambda],u) \in T^*M \times \mathbb{R} : [x,\lambda] \in T^*M \text{ and } u = \nu_b([x,\lambda]) = b\} \qquad (2.13b)$$

We now extend the definition of extremal surfaces so as to encompass the surfaces E^a and E^b as well, i.e., from now onwards by extremal surfaces we mean the surfaces E^a and E^b in addition to all the surfaces meant so far. It may be noted that the new set of extremal surfaces do not possess the property (i) of the old set.

An extremal surface

$$E = \{([x,\lambda],u) \in T^*M \times \mathbb{R} : [x,\lambda] \in S,\ u = \nu([x,\lambda])\}$$

is said to represent the optimal control at a point $[x,\lambda] \in T^*M$ if there exists a neighbourhood N of $[x,\lambda]$ such that the function $\nu([x,\lambda])$ represents the optimal control in N (see page 31). The definition implies that E represents the optimal control at every point in N.

We recall that Z is the set of all points in T^*M through which there exists an optimal path and assume that

$$\mathrm{Int}(Z) \neq \varphi \quad \text{and } Z \subseteq \overline{\mathrm{Int}\ (Z)}$$

Let Σ be an open set in T^*M such that $\Sigma \subseteq Z$ and for all points $[x,\lambda] \in \Sigma$,

(A_1) $H([x,\lambda],u)$ considered as a function of u has only non-degenerate critical points in the closed interval $[a,b]$ and

(A_2) the points $([x,\lambda],a) \in E^a$ and $([x,\lambda],b) \in E^b$ do not belong to any other extremal surface.

Choose two points $[x_1,\lambda_1]$ and $[x_2,\lambda_2]$ in Σ such that (i) the Hamiltonian $H([x_i,\lambda_i], u)$ considered as a function of u in $[a,b]$ has a single global maximum; $i = 1,2$, and (ii) there exists an optimal path $\alpha : [t_1,t_2] \to T^*M$ joining the two points and contained entirely in Σ. Let u(t) be the governing optimal control.

The way Σ is constructed enables us to invoke Lemma 1 and assert that around every point $[x,\lambda] \in \Sigma$, at which $H([x,\lambda],u)$ considered as a function of u in $[a,b]$ has a single global maximum, there exists a neighbourhood N and an extremal surface E which represents the optimal control in N. Let E^1 and E^2 be the extremal surfaces which represent the optimal control in suitable neighbourhoods of $[x_1,\lambda_1]$ and $[x_2,\lambda_2]$ respectively. Using the notation followed in (2.13) we write

$$E^i = \{([x,\lambda],u) \in T^*M \times \mathbb{R} : [x,\lambda] \in S_i \text{ and } u = \nu_i([x,\lambda])\};$$
$$i = 1,2.$$

We then have as a consequence

$$u(t_1) = \nu_1([x_1,\lambda_1]) \text{ and } u(t_2) = \nu_2([x_2,\lambda_2]).$$

If now we assume that no other extremal surface except for E^1 and E^2 takes over the role of representing the optimal control along the optimal path α, then there exists a point

$[x', \lambda']$ on α where the optimal control switches from E^1 to E^2.

By the statement "the optimal control switches from E^1 to E^2 at $[x', \lambda']$" we mean that at every point on the portion of the optimal path between $[x_1, \lambda_1]$ and $[x', \lambda']$ - the point $[x', \lambda']$ excluded - the optimal control is represented by the extremal surface E^1 or equivalently by the continuously differentiable function $\nu_1([x, \lambda])$ while at points on the portion between $[x', \lambda']$ and $[x_1, \lambda_1]$ - $[x', \lambda']$ excluded - it is represented by E^2 or equivalently by $\nu_2([x, \lambda])$.

As the extremal surfaces do not intersect it follows that

$$\nu_1([x', \lambda']) \neq \nu_2([x', \lambda']).$$

Therefore at $t' \in (t_1, t_2)$, the time when we reach the point $[x', \lambda']$ traversing the optimal path α, the governing optimal control $u(t)$ suffers a discontinuity. The value of the optimal control at $[x', \lambda']$ according to our convention (Section 1-1) is given by $\nu_2([x', \lambda'])$. We also note that if there are extremal surfaces other than E^1 and E^2 which take up the role of representing the optimal control between $[x_1, \lambda_1]$ and $[x_2, \lambda_2]$, then we simply have more number of switchings similar to the one just discussed.

We now proceed to analyse the reasons for which the optimal control switches from one extremal surface to another. For an extremal surface

$$E^i = \{([x,\lambda],u) \in T^*M \times \mathbb{R} : [x,\lambda] \in S_i \text{ and } u = \nu_i([x,\lambda])\}$$

where S_i and ν_i have their usual meaning, we define a vector field X_{ν_i} on S_i by the formula

$$X_{\nu_i}([x,\lambda]) = \sum_{k=1}^{n} f_k(x,\nu_i([x,\lambda]))\frac{\partial}{\partial x_k}$$

$$+ \sum_{k=1}^{n} \left(-\sum_{j=1}^{n} \lambda_j \frac{\partial f_j(x,\nu_i([x,\lambda]))}{\partial x_k}\right) \frac{\partial}{\partial \lambda_k} \qquad (2.14)$$

where $x_1, x_2, \ldots, x_n; \lambda_1, \lambda_2, \ldots, \lambda_n$ define a coordinate neighbourhood about the point $[x,\lambda] \in S_i$ in the sense that $x_1, x_2, \ldots, x_n$ define a coordinate neighbourhood about $x = \pi [x,\lambda] \in M$ and $\lambda = \sum_i \lambda_i \, dx_i$ (See page 24).

Let $[x_o,\lambda_o]$ be a point in Σ. For the sake of argument let us say that the Hamiltonian $H([x_o,\lambda_o],u)$ considered as a function of u in $[a,b]$ has a single global maximum. Let $E^1, E^2, \ldots, E^m$, be the m extremal surfaces (these include E^a and E^b also) such that each of the respective open sets $S_1, S_2, \ldots, S_m$ contains the point $[x_o,\lambda_o]$ (and the collection $\{E^i\}$, $i = 1,2,\ldots,m$, contains all such extremal surfaces). Of these let E^k represent the optimal control in a neighbourhood of $[x_o,\lambda_o]$. Then we must have

$$H([x_o,\lambda_o], \nu_k([x_o,\lambda_o])) > H([x_o,\lambda_o], \nu_i([x_o,\lambda_o]))$$

$$i = 1,2,\ldots,m, \; i \neq k.$$

The optimal path starting from $[x_o, \lambda_o]$ and in its immediate vicinity, is nothing but the integral curve of the vector field X_{ν_k} through $[x_o, \lambda_o]$. Now the rate of variation of the Hamiltonian function $H([x,\lambda], \nu_i([x,\lambda]))$ as we proceed along the integral curve of X_{ν_k} is given by the Lie derivative $L_{X_{\nu_k}} H([x,\lambda], \nu_i([x,\lambda]))$ [3, p.5-24]. Let $x_1, x_2, \ldots, x_n; \lambda_1, \lambda_2, \ldots, \lambda_n$ define a coordinate neighbourhood about the point $[x,\lambda]$, we then have

$$L_{X_{\nu_k}} H([x,\lambda], \nu_i([x,\lambda])) = \sum_{j=1}^{n} \frac{\partial H([x,\lambda], \nu_i([x,\lambda]))}{\partial x_j} \frac{dx_j}{dt}$$

$$+ \sum_{j=1}^{n} \frac{\partial H([x,\lambda], \nu_i([x,\lambda]))}{\partial \lambda_j} \frac{d\lambda_j}{dt}$$

$$+ \frac{\partial H([x,\lambda], \nu_i([x,\lambda]))}{\partial u} \left(\sum_{j=1}^{n} \frac{\partial \nu_i([x,\lambda])}{\partial x_j} \frac{dx_j}{dt} \right.$$

$$\left. + \sum_{j=1}^{n} \frac{\partial \nu_i([x,\lambda])}{\partial \lambda_j} \frac{d\lambda_j}{dt} \right).$$

Now $$\frac{\partial H([x,\lambda], \nu_i([x,\lambda]))}{\partial u} \equiv h([x,\lambda], \nu_i([x,\lambda])) \equiv 0$$

and $$\frac{dx_j}{dt} = \frac{\partial H([x,\lambda], \nu_k([x,\lambda]))}{\partial \lambda_j}$$

$$\frac{d\lambda_j}{dt} = - \frac{\partial H([x,\lambda], \nu_k([x,\lambda]))}{\partial x_j}$$

Hence $L_{X_{\nu_k}} H([x,\lambda], \nu_i([x,\lambda]))$

$$= \sum_{j=1}^{n} \frac{\partial H([x,\lambda],\ \nu_i([x,\lambda]))}{\partial x_j} \frac{\partial H([x,\lambda],\ \nu_k([x,\lambda]))}{\partial \lambda_j}$$

$$- \sum_{j=1}^{n} \frac{\partial H([x,\lambda],\ \nu_i([x,\lambda]))}{\partial \lambda_j} \frac{\partial H([x,\lambda],\ \nu_k([x,\lambda]))}{\partial x_j}$$

$$\Longrightarrow L_{X_{\nu_k}} H([x,\lambda],\ \nu_k([x,\lambda])) = 0.$$

In otherwords the Hamiltonian $H([x,\lambda],\ \nu_k([x,\lambda]))$ remains constant, at the value 1, say, as we proceed along the integral curve of X_{ν_k} while the other Hamiltonian functions $H([x,\lambda],\ \nu_i([x,\lambda]))$ vary along it.

At the point $[x_o,\lambda_o]$, $H([x,\lambda],\ \nu_k([x,\lambda]))$ is greater than all $H([x,\lambda],\ \nu_i([x,\lambda]))$, however, as we proceed along the optimal path through $[x_o,\lambda_o]$ some of the Hamiltonian functions $H([x,\lambda],\ \nu_i([x,\lambda]))$ may start increasing and catch up with the value 1. Let $[x',\lambda'] \varepsilon\ \Sigma$ be the earliest point on the optimal path when this happens - where the function $H([x,\lambda],\ \nu_j([x,\lambda]))$, (where j is an integer between 1 and m, $j \neq k$) attains the value 1. The Hamiltonian $H([x,\lambda],u)$ considered as a function of u in $[a,b]$ therefore has two global maxima when evaluated with $[x,\lambda] = [x',\lambda']$

$$1 = H([x',\lambda'],\ \nu_k([x',\lambda'])) = H([x',\lambda'],\ \nu_j([x',\lambda']))$$

$$> H([x',\lambda'],\ \nu_i([x',\lambda'])); \ i \neq k,\ i \neq j.$$

$[x',\lambda']$ is a potential point where the optimal control may

switch from E^k to E^j. We discuss this below, with regard to which we take note of the fact that

$$L_{X_{\nu_k}} H([x,\lambda], \nu_j([x,\lambda])) = -L_{X_{\nu_j}} H([x,\lambda], \nu_k([x,\lambda])).$$

(i) Consider first the case where at the point $[x',\lambda']$ we have

$$L_{X_{\nu_j}} H([x,\lambda], \nu_k([x,\lambda])) < 0. \tag{2.15}$$

This implies that $L_{X_{\nu_k}} H([x,\lambda], \nu_j([x,\lambda])) > 0$ at $[x,\lambda] = [x',\lambda']$. If from the point $[x',\lambda']$ we traverse the integral curve of X_{ν_k}, $H([x,\lambda], \nu_j([x,\lambda]))$ increases while $H([x,\lambda], \nu_k([x,\lambda]))$ remains constant. Hence at all points $[x,\lambda]$ on this integral curve, in the immediate vicinity of $[x',\lambda']$ we have

$$H([x,\lambda], \nu_j([x,\lambda])) > H([x,\lambda], \nu_k([x,\lambda])) = 1 >$$
$$H([x,\lambda], \nu_i([x,\lambda])); \; i \neq k, \; i \neq j.$$

On the other hand if we go along the integral curve of X_{ν_j} from $[x',\lambda']$ instead, $H([x,\lambda], \nu_k([x,\lambda]))$ decreases while $H([x,\lambda], \nu_j([x,\lambda]))$ remains constant at the value 1. At points $[x,\lambda]$ on this integral curve close to the point $[x',\lambda']$ we then have

$$1 = H([x,\lambda], \nu_j([x,\lambda])) > H([x,\lambda], \nu_i([x,\lambda])), \; i \neq j.$$

Maximum principle therefore dictates that the optimal path from $[x',\lambda']$ onwards and close to it will be the integral curve of the vector field X_{ν_j} through $[x',\lambda']$. Thus the

optimal control switches from extremal surface E^k to the extremal surface E^j at $[x', \lambda']$.

(ii) If on the other hand at the point $[x', \lambda']$ we have

$$L_{X_{\nu_j}} H([x,\lambda], \nu_k([x,\lambda])) > 0 \tag{2.16}$$

a similar argument leads us to conclude that the optimal control does not switch at $[x', \lambda']$. The optimal path from $[x', \lambda']$ onwards continues to be along the integral curve of X_{ν_k}.

(iii) If we have instead

$$L_{X_{\nu_k}} H([x',\lambda'], \nu_j([x',\lambda'])) = 0 \text{ and}$$

$$L_{X_{\nu_j}} H([x',\lambda'], \nu_k([x',\lambda'])) = 0 \tag{2.17}$$

it may happen that $H([x,\lambda], \nu_k([x,\lambda]))$ remains constant along the integral curve of X_{ν_j} through $[x', \lambda']$ while $H([x,\lambda], \nu_j([x,\lambda]))$ remains constant along the integral curve of X_{ν_k} through the same point. Then as far as maximum principl goes we have equal liberty of choosing either of these integral curves as our optimal path. We encounter a singular problem here. Such situations will not arise as by assumption our system is non-singular. We therefore expect the Hamiltonian $H([x,\lambda], \nu_k([x,\lambda]))$ $\left(H([x,\lambda], \nu_j([x,\lambda]))\right)$ to vary along the integral curve of X_{ν_j} (X_{ν_k}) through $[x', \lambda']$. The nature of this variation can be determined by examining the higher order

Lie derivatives of $H([x,\lambda], \nu_k([x,\lambda]))$ $(H([x,\lambda], \nu_j([x,\lambda])))$ with respect to X_{ν_j} (X_{ν_k}). An application of reasoning similar to that developed in (i) and (ii) above can then determine the optimal path through $[x',\lambda']$.

When the Hamiltonian $H([x',\lambda'], u)$ considered as a function of u in $[a,b]$ has more than two global maxima, the optimal path from $[x',\lambda']$ onwards can be chosen by considerations essentially similar to those in (i), (ii) and (iii) above For example if at the point $[x',\lambda']$ we have

$$1 = H([x',\lambda'], \nu_k([x',\lambda'])) = H([x',\lambda'], \nu_j([x',\lambda']))$$
$$= H([x',\lambda'], \nu_p([x',\lambda'])) > H([x',\lambda'], \nu_i([x',\lambda']))$$
$$i \neq j,k,p$$

then

$$L_{X_{\nu_k}} H([x',\lambda'], \nu_j([x',\lambda'])) < 0 \text{ and}$$

$$L_{X_{\nu_k}} H([x',\lambda'], \nu_p([x',\lambda'])) < 0$$

imply that the optimal path from $[x',\lambda']$ onwards will be along the integral curve of X_{ν_k} through $[x',\lambda']$. The equalities

$$L_{X_{\nu_k}} H([x',\lambda'], \nu_j([x',\lambda'])) = L_{X_{\nu_j}} H([x',\lambda'], \nu_p([x',\lambda']))$$
$$= L_{X_{\nu_p}} H([x',\lambda'], \nu_k([x',\lambda'])) = 0$$

on the other hand demand the examination of higher order Lie derivatives to determine the optimal path through $[x',\lambda']$.

An altogether new situation arises if the inequalities

$$L_{X_{\nu_k}} H([x',\lambda'],\ \nu_j([x',\lambda'])) > 0,$$

$$L_{X_{\nu_j}} H([x',\lambda'],\ \nu_p([x',\lambda'])) > 0 \text{ and}$$

$$L_{X_{\nu_p}} H([x',\lambda'],\ \nu_k([x',\lambda'])) > 0 \text{ hold.}$$

Evidently from the point $[x',\lambda']$ onwards the optimal path **cannot** be along the integral curve of any one of the vector fields X_{ν_j}, X_{ν_k}, X_{ν_p}. However, in a well defined system the optimal path ought to be able to continue past $[x',\lambda']$. The way out is, for the Hamiltonian $H([x',\lambda'], u)$ considered as a function of u in $[a,b]$ to have developed more than three global maxima.

In general therefore all points $[x',\lambda'] \varepsilon \Sigma$ where the Hamiltonian $H([x',\lambda'], u)$ considered as a function of u in $[a,b]$ has multiple global maxima, are potential switching points of the optimal control and the mechanism of switching is merely a generalized version of the process just discussed.

The analysis, so far, leads us to several important facts. First of all we note that the open set Σ contains two types of points :

(τ_1) The points $[x,\lambda]$ for which the Hamiltonian $H([x,\lambda],u)$ considered as a function of u in $[a,b]$ has a single global maximum and

(τ_2) The points for which it has multiple global maxima.

From Lemma 1 it follows that the set of points of the type (τ_1) is an open subset of Σ. The assumption that our optimal control is non-singular with respect to the maximum principle prohibits the existence of a non-empty interior for the set of points of the type (τ_2). For, suppose there exists a point $[x',\lambda'] \varepsilon \Sigma$ and a neighbourhood N around it such that N consists of points of the type (τ_2) only, then the optimal path through $[x',\lambda']$ penetrates N. However, once we are inside N, maximum principle fails to specify the control and it will not be possible to construct the optimal path using it.

At every point of the type (τ_1) the optimal control is represented by one of the extremal surfaces, while at points of the type (τ_2) the optimal control possibly switches from one extremal surface to another. Let A^i denote the set of points in Σ at which E^i represents the optimal control and we consider all such A^i (corresponding to all extremal surfaces E^i) which are non-empty. A^i are evidently open subsets of Σ and we have

$$A^i \cap A^j = \varphi \text{ for } i \neq j.$$

We decompose each A^i into its components

$$A^i = \bigcup_k a^i_k$$

where each a^i_k is a maximally connected component of A^i. Clearly the subsets a^i_k (for all i and all k) are open subsets of Σ

disjoint from each other. In the relative topology of Σ the boundary of each a_k^i consists of points of the type (τ_2) where the optimal control switches from E^i to other extremal surfaces or vice versa. There may also exist points on the boundary such that for an optimal path through it the optimal control does not suffer any switchings there (see condition (2.16)). Moreover it is evident that

$$\Sigma = \bigcup_i \bar{A}_i = \bigcup_i \left(\bigcup_k \bar{a}_k^i \right)$$

where the closure is taken in the relative topology of Σ. The optimal control at every point in a_k^i and at some points on its boundary is represented by the continuously differentiable function $\nu_i([x,\lambda])$. Let C_{i_k} denote the set of all such points, we have

$$C_{i_k} \subseteq \bar{a}_k^i$$

and Σ is the disjoint union

$$\Sigma = \bigcup_{i,k} C_{i_k} .$$

This enables us to conclude the following :

There exists a partition of Σ by partition sets $\{C_i\}$, say, such that

(i) C_i have a non-empty interior and

$$C_i \subseteq \overline{\mathrm{Int}(C_i)} ,$$

(ii) C_i are connected,

(iii) there exists a continuously differentiable function $\mu_i : S_i \to \mathbb{R}$ — S_i is an open subset of Σ with $C_i \subseteq S_i$ — such that μ_i represents the optimal control on C_i,

(iv) C_i are maximal with respect to (ii) and (iii).

(Note : Closures and interiors are with respect to the relativ topology of Σ).

The partition sets C_i will be called control regions. In the next section we generalize our analysis to construct control regions for the whole of Z.

Remark 1. When an optimal path crosses over from one control region to another the optimal control switches from one extrem surface to another and suffers a discontinuity in the process. We reserve the term "discontinuous switching" to describe this phenomenon. This is because at a later stage we encounter a phenomenon where the optimal control switches from one extrema surface to another without suffering discontinuities.

2-3 Control Regions

To effect a partition of Z on the lines suggested in last section, we analyse the two assumptions which were used to define Σ (See page 40). For, if we succeed in disengaging ourselves from those two assumptions and still be able to construct control regions, our purpose would have been served.

The first assumption for Σ stipulates that for the points $[x,\lambda] \in \Sigma$, $H([x,\lambda],u)$ considered as a function of u in $[a,b]$ has only non-degenerate critical points. If we neg

this and allow degenerate critical points to occur as well, a host of new situations arise. This is because if $\nu \in [a,b]$ is a degenerate critical point, through the point $([x,\lambda],\nu) \in T^{*}M \times \mathbb{R}$, we can no more expect a unique local extremal surface for the point $[x,\lambda]$ to exist. In fact it is possible to have more than one local extremal surfaces or none at all through $([x,\lambda],\nu)$. This follows from the theory of branching of solutions [6].

From now on, we abbreviate the statement "critical points of $H([x,\lambda],u)$ considered as a function of u in $[a,b]$" to read "critical points of $H([x,\lambda],u)$ in $[a,b]$". For a given $[x,\lambda]$, if $H([x,\lambda],u)$ has only isolated critical points in $[a,b]$, it follows that they are finite in number. On the other hand if the critical points in $[a,b]$ are not all isolated, it is possible to have an infinite number of isolated critical points in $[a,b]$. This however is an extremely unusual situation and very difficult to treat. To obviate such difficulties we make the following assumption. We assume that $H([x,\lambda],u)$ can have critical points in $[a,b]$ in either or both of the following two ways :

(i) it can have finitely many isolated critical points,

(ii) there may exist an interval $[c,d]$, $a \leq c < d \leq b$, such that every point in $[c,d]$ is a critical point. In such cases $H([x,\lambda],u)$ considered as a function of u remains constant in $[c,d]$.

We further categorize these critical points into the followin
three categories :

(C1) From among the critical points of $H([x,\lambda],u)$ in $[a,b]$ we first choose those points, say, $u_1, u_2, \ldots, u_m$ lying in $(a,b$ such that through each $([x,\lambda],u_i) \in T^*M \times \mathbb{R}$; $i = 1,2,\ldots,m$, there exists a unique local extremal surface for the point $[x,\lambda]$. The set $\{u_1, u_2, \ldots, u_m\}$ therefore consists of all non-degenerate critical points together with some well behaved degenerate critical points, and has a finite number of elemen because it is a subset of the set consisting of all isolated critical points in $[a,b]$. We now extend the above selection to all points in T^*M, construct local extremal surfaces (when possible) for each of them and join the overlapping local extremal surfaces. We end up as before with a number of $2n$-dimensional submanifolds in $T^*M \times \mathbb{R}$; these are our new "extremal surfaces". As before we stipulate that the extrema surfaces are maximally connected. We also construct the surfaces E^a and E^b as defined earlier and call them extremal surfaces. If $u = a(b)$ happens to be a critical point of $H([x,\lambda],u)$ for some $[x,\lambda] \in T^*M$, it is taken care of by the extremal surfaces $E^a(E^b)$. In the process we also allow $E^a(E^b$ to have points in common with other extremal surfaces, thereb negating the second assumption required to select Σ (See page

(C2) Next consider the degenerate critical points of $H([x,\lambda],u)$ of the type $\eta \in [a,b]$, such that the point $([x,\lambda],\eta)$ belongs to the boundary of some extremal surface E

as constructed in (C1)- and such that it is possible to extend E smoothly up to the point $([x,\lambda],\eta)$ (Here by the boundary of E we mean the set $\bar{\bar{E}} - E$). We extend E smoothly up to the point $([x,\lambda],\eta)$ and continue doing this for all $[x,\lambda] \varepsilon T^*M$ for which $H([x,\lambda],u)$ has critical points of the type η.

Now critical points of the type η are not of the category (C1), however, for some of them it may be possible to have more than one local extremal surfaces through $([x,\lambda],\eta)$. In such cases we can extend the extremal surfaces (as constructed in (C1)), which overlap with the local extremal surfaces through $([x,\lambda],\eta)$, up to the point $([x,\lambda],\eta)$, by the simple process of joining them with the local extremal surfaces they overlap with. We now extend the definition of the extremal surfaces to encompass the extended extremal surfaces as well. It is now possible for any two extremal surfaces to have common points and these common points (for surfaces not involving E^a and E^b) are of the type $([x,\lambda],\eta)$ just discussed.

(C3) In the third category we dump all degenerate critical points of $H([x,\lambda],u)$ which are neither of the type (C1) nor (C2). Let $\eta' \varepsilon [a,b]$ denote such a critical point. If G be the set of all points in a neighbourhood of $([x,\lambda],\eta')$ at which $h([x,\lambda],u) = 0$, G can have rather bizarre shape.

We assume that for an optimal path through $[x,\lambda]$, η' is barred from representing the value of the optimal control at $[x,\lambda]$ by the non-singularity condition. That this is true for

well behaved systems of the form (2.1) can be seen directly from the following discussion.

First of all we note that an isolated critical point of $H([x,\lambda],u)$ is either of the type (C1) or (C2) and therefore not of the type (C3). On the other hand let all the points in some interval $[c,d]$, $a \leq c < d \leq b$, be critical points of $H([x,\lambda],u)$. $H([x,\lambda],u)$ considered as a function of u remains constant in $[c,d]$. If $\eta' \in [c,d]$ is not of the type (C2) then it is of the type (C3). Now if $H([x,\lambda],u)$ considered as a function of u in $[a,b]$ attains global maximum at $u = \eta'$, it does so at all points $u \in [c,d]$; with regard to this property therefore it is not possible to distinguish η' from other points in $[c,d]$. In other words the fact that η' represents the value of the optimal control at $[x,\lambda]$ cannot be concluded from the maximum principle. But the optimal control is assumed to be non-singular it is therefore clear that for a possible optimal path through $[x,\lambda]$ the value of the optimal control there cannot be η'.

We can therefore forget the critical points of the type (C3). Critical points of the types (C1) and (C2), only, can represent the value of the optimal control and they lie in one or other of the extremal surfaces.

This concludes the discussion of situations that arise when the two assumptions that were required to select Σ are ignored.

The extremal surfaces, as redefined, retain their old form with a slight change. If E^i is an extremal surface we can write

$$E^i = \{([x,\lambda],u) \in T^*M \times \mathbb{R} : [x,\lambda] \in S_i \text{ and } u = \nu_i([x,\lambda])\}.$$

Here $S_i = \Psi(E^i)$ need no more be open always. ν_i, however as of old, is a continuously differentiable function defined on S_i (which is possible by virtue of the construction of E^i, see alpage 29) such that $h([x,\lambda],\nu_i) = 0 \;\forall\; [x,\lambda] \in S_i$. Further

$$E^a = \{([x,\lambda],u) \in T^*M \times \mathbb{R} : [x,\lambda] \in T^*M \text{ and } u = \nu_a([x,\lambda]) = a\}$$

$$E^b = \{([x,\lambda],u) \in T^*M \times \mathbb{R} : [x,\lambda] \in T^*M \text{ and } u = \nu_b([x,\lambda]) = b\}$$

We now turn to the problem of partitioning Z into control regions. Let the optimal control in a neighbourhood of a point $[x_1,\lambda_1] \in Z$ be represented by the extremal surface E^1. The optimal path through $[x_1,\lambda_1]$ and in its immediate vicinity is along the integral curve of the vector field X_{ν_1}. Traversi this optimal path, we reach a point $[x^*,\lambda^*]$, say, where the value of the optimal control is u^* and suppose $([x^*,\lambda^*],u^*)$ is common to the extremal surfaces $E^1,E^2,\ldots,E^m$. The optimal path from $[x^*,\lambda^*]$ onwards is along the integral curve of one of the vector fields $X_{\nu_1},X_{\nu_2},\ldots,X_{\nu_m}$ and the appropriate vector field can be determined by an analysis similar to that employed in the last section to study the discontinuous switching of the optimal control. For example the optimal pa

is along the integral curve of X_{ν_2} if the inequality

$$L_{X_{\nu_2}} H([x^*,\lambda^*], \nu_i([x^*,\lambda^*])) < 0 \tag{2.18}$$

holds for all $i \in \{1,3,\ldots,m\}$ for which the Hamiltonian $H([x,\lambda], \nu_i([x,\lambda]))$ is defined along the integral curve of X_{ν_1} through the point $[x^*,\lambda^*]$. We say that the optimal control switches continuously from E^1 to E^2 at $[x^*,\lambda^*]$. This is because at $[x^*,\lambda^*]$

$$\nu_1([x^*,\lambda^*]) = \nu_2([x^*,\lambda^*]).$$

We call switchings of this type "continuous switchings" to emphasize the fact that they are different in nature from the discontinuous switchings discussed in Section (2-2). While discontinuous switchings make the optimal control discontinuou[s] continuous switchings make it non-differentiable at the worst.

With regard to condition (2.18) we would like to mention a special case. Let $H([x^*,\lambda^*],u)$ considered as a function of u in $[a,b]$ have a unique global maximum at $u = b$ and let $([x^*,\lambda^*],b)$ be common to both E^b and E^i. The optimal path from $[x^*,\lambda^*]$ will not be along X_{ν_i} if $L_{X_{\nu_i}} \nu_i([x,\lambda]) > 0$ at $[x^*,\lambda^*]$. This is so even if the supporting inequality

$$L_{X_{\nu_i}} H([x,\lambda], \nu_b([x,\lambda])) < 0$$

holds. For, otherwise, the value of the optimal control will exceed b which is not admissible. The optimal path will be

along the integral curve of X_{ν_b} through the point $[x^*, \lambda^*]$.

The above analysis shows that there exist the following three types of points in Z.

(T_1) Points of the type $[x, \lambda] \varepsilon$ Z at which the optimal control is represented by one of the extremal surfaces

(T_2) Points of the type $[x', \lambda'] \varepsilon$ Z such that as we traverse the optimal path through $[x', \lambda']$ the optimal control switches discontinuously from one extremal surface to another at this point.

(T_3) Furthermore points of the type $[x^*, \lambda^*]$ such that as we traverse the optimal path through $[x^*, \lambda^*]$ the optimal control switches continuously from one extrema surface to another at this point.

The analysis, from now on, is confined to the point in the subset $Z \subseteq T^*M$ - the topology therefore is the relativ topology of Z.

The set of points of the type (T_1) is an open subset of while the set of points of the type (T_2) has an empty interio this was established in the last section. Furthermore, it is clear from the definition of the extremal surfaces - as redef in this section - that the set of points of the type (T_3), to has an empty interior.

Let A^i denote the set of all points in Z at which E^i represents the optimal control and we consider all such A^i (corresponding to all extremal surfaces E^i) which are non-em

A^i are evidently open subsets of Z and we have

$$A^i \cap A^j = \varphi \text{ for } i \neq j.$$

We decompose each A^i into its components

$$A^i = \bigcup_k a^i_k$$

where each a^i_k is a maximally connected component of A^i. Clearl the subsets a^i_k are open subsets of Z disjoint from each other. The boundary of each a^i_k consists of points of the types (T_2) and (T_3) where the optimal control switches discontinuously or continuously from E^i to other extremal surfaces or vice ve There may also exist points on ∂a^i_k such that for an optimal path through it the optimal control does not suffer any switchings there (See condition (2'.16)). Moreover it is evident that

$$Z = \bigcup_i \bar{A}_i = \bigcup_i \left(\bigcup_k \bar{a}^i_k\right)$$

The optimal control at every point in a^i_k and at some points on its boundary is represented by the continuously differenti function $\nu_i([x,\lambda])$. Let C_{i_k} denote the set of all poin then

$$C_{i_k} \subseteq \bar{a}^i_k$$

and Z is the disjoint union

$$Z = \bigcup_{i,k} C_{i_k}.$$

This enables us to conclude the following :

There exists a partition of Z by the partition sets $\{C_i\}$, say, such that :

(i) C_i have a non-empty interior and

$$C_i \subsetneq \overline{\mathrm{Int}(\overline{C}_i)}$$

(ii) C_i are connected

(iii) there exists a continuously differentiable function $\mu_i : S_i \to \mathbb{R}$ - with $C_i \subseteq S_i$ - such that μ_i represents the optimal control in C_i and

(iv) C_i are maximal with respect to (ii) and (iii).

The partition sets $\{C_i\}$ will be called control region The family of continuously differentiable functions $\{\mu_i\}$ collectively define the optimal control in the form

$$u : Z \to \mathbb{R}\,; \tag{*}$$

u being defined by the straightforward rule

$$u([x,\lambda]) = \mu_i([x,\lambda]) \text{ whenever } [x,\lambda] \in C_i\,.$$

When an optimal path crosses over from one control region to another, the optimal control when represented in the form (*) suffers a discontinuous or a continuous switching.

Let the optimal path through the point $[x_0,\lambda_0]$ be governed by the optimal control, when expressed as a function of time, u(t). We start from $[x_0,\lambda_0]$ at time $t = 0$ and reach $[x_1,\lambda_1]$ (say) at time $t = \pi$ where the optimal control when represented in the form (*) switches discontinuously

from one extremal surface to another, it follows that at $t = \tau$ $u(t)$ suffers a discontinuity. On the other hand if the optima control, when represented in the form (*), switches continuous at $[x_1, \lambda_1]$ but becomes non-differentiable in the process, $u(t)$ is likely to be non-differentiable at $t = \tau$. Conversely it is clear that as we traverse the optimal path a discontinui or a non-differentiability of $u(t)$ occurs only when the optima control when represented in the form (*) switches from one extremal surface to another.

We conclude this chapter with some observations regardi the nature of the control regions. Let C_i and C_j be two control regions in Z. We denote by C_{ij} the set of all points in ∂C_i where the optimal path crosses over from the control region C_i into the control region C_j. Below, we make some comments about these sets C_{ij} for well behaved systems.

We consider a simple case to begin with. Let C_1 and C_2 be two control regions and let the optimal control in $Int(C_1)$ and $Int(C_2)$ be represented by the extremal surfaces E^1 and E^2 respectively. Further, suppose, the control region C_1 is such that there exist optimal paths from C_1 to C_2 or from C_2 to C_1 or both and that no other control regions except for C_2 is similarly related to C_1.

At all points $[x,\lambda] \varepsilon C_{12}$ we have

$$g([x,\lambda]) = H([x,\lambda], \nu_1([x,\lambda])) - H([x,\lambda], \nu_2([x,\lambda])) = 0$$

and

$$L_{X_{\nu_1}} H([x,\lambda] , \nu_2([x,\lambda])) \geq 0 . \qquad (\theta_1)$$

Similarly at all points $[x,\lambda] \varepsilon C_{21}$ we have

$$g([x,\lambda]) = H([x,\lambda], \nu_1([x,\lambda])) - H([x,\lambda], \nu_2([x,\lambda])) = 0$$

and

$$L_{X_{\nu_2}} H([x,\lambda] , \nu_1([x,\lambda])) \geq 0 . \qquad (\theta_2)$$

Now $g([x,\lambda])$ is a smooth map and has rank 1 at every point where either the condition (θ_1) or the condition θ_2, without the equality sign, is satisfied. In well behaved cases the situation therefore can be as follows :

(i) Both Int (C_{12}) and Int (C_{21}) (interior in the relative topology of ∂C_1) are (2n-1)-dimensional smooth manifol conditions (θ_1) and (θ_2) without equality sign being satisfied respectively.

(ii) Int (C_{12}) and Int (C_{21}) have a common boundary Δ (in tl relative topology of ∂C_1) and ∂C_1 is the disjoint union

$$\partial C_1 = \text{Int } (C_{12}) \cup \text{Int } (C_{21}) \cup \Delta .$$

(iii) At points $[x,\lambda] \varepsilon \Delta$

$$g([x,\lambda]) = 0$$

and

$$L_{X_{\nu_1}} H([x,\lambda] , \nu_2([x,\lambda])) = 0.$$

Now whether the optimal path crosses over from C_1 to C_2 or from C_2 to C_1 at a point, say, $[x',\lambda'] \varepsilon \Delta$ can be determined by examining the higher order Lie derivatives as mentioned before. In most cases the optimal path turns back into the control region from which it came, at points (**not all**) belon to Δ - we will see this later.

In general given the control region C_i we can find more than one other control regions of the type C_j such that there exist optimal paths from C_i to C_j or from C_j to C_i or both. The sets C_{ij}, C_{ji} are subsets of ∂C_i. Consider all such subsets which are non-empty. In well behaved systems $\mathrm{Int}(C_{ij})$, $\mathrm{Int}(C_{ji})$ (interior in the relative topology of ∂C_i) are $(2n-1)$ dimensional smooth manifolds and $C_{ij} \subseteq \overline{\mathrm{Int}(C_{ij})}$, similarly $C_{ji} \subseteq \overline{\mathrm{Int}(C_{ji})}$. $\overline{C_{ij}}$, $\overline{C_{ji}}$ are therefore $(2n-1)$ dimensional manifolds with boundary. These boundaries are not smooth in general.

Let Δ denote the set of points such that $[x^*,\lambda^*] \varepsilon \Delta \Longrightarrow [x^*,\lambda^*]$ is a point common to the boundaries - in the relative topology of ∂C_i - of each of the sets $C_{ij_1}, C_{ij_2},\ldots,C_{ij_k}; C_{j_{k+1}i}, C_{j_{k+2}i},\ldots,C_{j_{k+m}i}$ and none else; $(k+m) \leq 2n$. Then Δ in well behaved cases is a smooth manifol without boundary, $2n-(k+m) \leq \dim \Delta < 2n-1$ and $\bar{\Delta} - \Delta$ consists of points which are common not only to the boundaries of $C_{ij_1},\ldots,C_{ij_k}; C_{j_{k+1}i},\ldots,C_{j_{k+m}i}$, but also to the boundaries of some other such sets.

Thus far we have been considering well behaved cases only. Given two control regions C_i and C_j such that there exist optimal paths from C_i into C_j, the set $\bar{C}_{ij}$ need not always be a (2n-1)-dimensional manifold. It can be a manifold (with boundary of course) with lesser dimensionality and in worst cases can even fail to qualify as a manifold.

Finally, we observe that if there exists an optimal path through $[x_o, \lambda_o]$ and if u_o is the value of the optimal control there, then there also exists an optimal path through $[x_o, \alpha \lambda_o]$, for all real $\alpha > 0$, where the optimal control has the value u_o again. Therefore for a control region C_i

$$[x, \lambda] \in C_i \Longrightarrow [x, \alpha\lambda] \in C_i \ \forall \ \alpha > 0.$$

With this we conclude the present analysis.

CHAPTER 3

TRANSVERSALITY CONDITION

3-1 Preliminary Remarks.

With the help of the constructions made in the last chapter we now proceed to seek conditions which when imposed on the system, under consideration, put an upper bound on the number of discontinuities of the optimal control. To facilitate this investigation we first define a few terms.

Let A be a submanifold of T^*M. We denote by $\partial^M A$ the boundary, in the manifold sense, of the submanifold A and by ∂A the topological boundary of the set A. Further $\mathrm{Int}(A)=A-\partial A$ and $\mathrm{Int}^M(A) = A-\partial^M A$, [4, pp. 57-58].

Let V and W be two boundaryless smooth submanifolds of a smooth manifold N. Then V is said to be transversal to W - denoted by $V \pitchfork W$ - either if

$$V \cap W = \varphi$$

or if $V \cap W \neq \varphi$ and for every $x \in V \cap W$

$$V_x + W_x = N_x.$$

These conditions are symmetric with respect to V and W hence $V \pitchfork W \iff W \pitchfork V$ [4, p.29].

3-2 The Transversality Condition.

We are considering the time optimal control problem as defined in Section (1-1). M, as we have said before, denotes

the state space and is of dimension n and T^*M denotes its cotangent bundle. We partition $Z \subseteq T^*M$ into the family $\{C_i\}$ of control regions.

(Note : The control regions are connected subsets of Z. There may however exist control regions of the type C_α such that $Int(C_\alpha)$ is not connected. For such control regions we adopt the following convention. We express $Int(C_\alpha)$ as the disjoint union of its components (maximally connected), say, in the form, $Int(C_\alpha) = D_1 \cup D_2 \cup \cdots \cup D_k$. Next we partition C_α by expressing it as the disjoint union, $C_\alpha = C_{\alpha_1} \cup C_{\alpha_2} \cup \cdots \cup C_{\alpha_k}$, such that $Int(C_{\alpha_s}) = D_s$, $1 \leq s \leq k$. We now treat each C_{α_s} as a separate control region). Let $\mu_i : C_i \to [a,b]$ be the continuously differentiable function that represents the optimal control everywhere on C_i. The function μ_i is actually defined on a bigger set $S_i \supseteq C_i$ which is the projection under Ψ of the extremal surface that represents the optimal control in $Int(C_i)$. We define a vector field X_i on S_i by the rule

$$X_i = \sum_{j=1}^{n} f_j(x,\mu_i([x,\lambda])) \frac{\partial}{\partial x_j} +$$

$$+ \sum_{j=1}^{n} \left(- \sum_{k=1}^{n} \lambda_k \frac{\partial f_k(x,\mu_i([x,\lambda]))}{\partial x_j}\right) \frac{\partial}{\partial \lambda_j} . \qquad (3.1)$$

Here $x_1, x_2, \ldots, x_n; \lambda_1, \lambda_2, \ldots, \lambda_n$ define a co-ordinate neighbourhood (See page 42) about the point $[x,\lambda]$ in T^*M. We denote by $\Phi_i(t,\cdot)$ the one parameter group of diffeomorphisms induced by X_i.

Let x_o be a point in the state space M. From among all the optimal trajectories that pass through x_o, we propose to choose one arbitrarily and count the number of discontinuitic of the optimal control along this trajectory. Equivalently we can choose a point $[x_o,\lambda_o] \varepsilon M^*_{x_o} \cap Z$ arbitrarily ($[x_o,\lambda_o] \neq [x_o,0]$) and study the number of switchings that the optimal control suffers as we traverse the optimal path through $[x_o,\lambda_o]$. In order to put a bound on the number of discontinuities of the optimal control it is sufficient to consider the discontinuous switchings only, however, it turns out that the continuous switchings have also a role to play. Let $C_{ij} \subseteq \partial C_i$ denote the subset of all points where the optimal path crosses over from the control region C_i to the control region C_j. We assume that $\mathrm{Int}(C_{ij})$-interior in the relative topology of ∂C_i - is a (2n-1)-dimensional smooth manifold and $C_{ij} \subseteq \overline{\mathrm{Int}(C_{ij})}$, this implies that $\bar{C}_{ij}$ is a (2n-1)-dimensional manifold with boundary (we do not exclude the possibility of this boundary being the null set). Situations where this assumption does not hold will be considered later. We propose to carry on the analysis in the following three steps :

(1) First we suppose that every time the optimal path from $[x_o,\lambda_o]$ crosses over from a control region, say, C_i to a control region, say, C_j it does so at a point belonging to $\mathrm{Int}^M(\bar{C}_{ij})$ and the optimal control suffers a discontinuous switching there.

(2) In the second step we allow the optimal control to switch in any manner, the optimal path however still continues to pass through $\mathrm{Int}^M(\bar{C}_{ij})$.

(3) In the third step we take off all restrictions, i.e., the optimal path is allowed to pass through $C_{ij} - \mathrm{Int}^M(\bar{C}_{ij})$ as well and the optimal control switches in all possible manners.

Step 1

Let $[x_o, \lambda_o]$ be an interior point of the control region C_1 and let $\Sigma = C_1 \cap M^*_{x_o}$. It is then possible to find a neighbourhood σ_o of $[x_o, \lambda_o]$ in the relative topology of $M^*_{x_o}$ such that $\sigma_o \subseteq \Sigma$. We choose a σ_o, sufficiently small, so that its size does not hinder the arguments that follow. σ_o is evidently a submanifold of T^*M having co-dimension n (all co-dimensions in this discussion are with respect to T^*M). The optimal path from a point in σ_o is along the integral curve of the vector field X_1 through that point. Since $\Phi_1(t,\cdot)$ is a diffeomorphism, it follows that $\Phi_1(t,\sigma_o)$ is a submanifold with co-dimension n for all t.

Let the optimal path from $[x_o, \lambda_o]$ meet the boundary ∂C_1 in time $t_1 > 0$. If no such t_1 exists it follows that the optimal control, as we traverse the optimal path from $[x_o, \lambda_o]$ onwards, suffers no switchings. This immediately implies that along the optimal trajectory, from $x_o \in M$, which arises as a projection of this optimal path, the optimal control suffers no discontinuity. This is a trivial case of

our analysis. We therefore assume the existence of a finite $t_1 > 0$. Let

$$[x_1, \lambda_1] = \Phi_1(t_1, [x_o, \lambda_o]) \in \partial C_1 .$$

The optimal path crosses over from the control region C_1 into the control region C_2 (say) at $[x_1, \lambda_1]$. Our assumptions imply that the optimal control switches discontinuously here and $[x_1, \lambda_1] \in \mathrm{Int}^M(\bar{C}_{12})$. Further, it is trivial to note that $[x_1, \lambda_1] \not\sim [x_1, 0]$. We now impose the crucial condition :

The System (1.1) is such that

$$\Phi_1(t, \sigma_o) \not\pitchfork \mathrm{Int}^M(\bar{C}_{12}) \quad \forall\, t > 0 . \tag{3.2}$$

Consider the intersection

$$\sigma_1 = \Phi_1(t_1, \sigma_o) \cap \mathrm{Int}^M(\bar{C}_{12}),$$

σ_1 is non-empty as $[x_1, \lambda_1] \in \sigma_1$ and it follows from (3.2) that it is a submanifold of T^*M with

$$\begin{aligned} \mathrm{codim}\ \sigma_1 &= \mathrm{codim}\ \Phi_1(t_1, \sigma_o) + \mathrm{codim}\ \mathrm{Int}^M(\bar{C}_{12}) \\ &= n+1 \end{aligned}$$

From the nature of the vector fields X_i it follows that if $[\tilde{x}, \tilde{\lambda}] = \Phi_i(t, [x, \lambda])$ then $[\tilde{x}, \alpha\tilde{\lambda}] = \Phi_i(t, [x, \alpha\lambda])$ for all real α (for, the co-state equations are linear in the co-state variable and $\mu_i([x, \lambda]) \equiv \mu_i([x, \alpha\lambda])$). Moreover, from the nature of the Hamiltonian it follows that if the optimal path through $[x, \lambda]$ crosses over from the control region C_i to the control region C_j at $[x, \lambda]$ then the optimal

path through $[x,\alpha\lambda]$ does the same at $[x,\alpha\lambda]$ for all real $\alpha > 0$.

Now as $\sigma_o \subset M^*_{x_o}$ is open in the relative topology of $M^*_{x_o}$, it follows that for all $[x,\lambda] \varepsilon\, \sigma_o$, the points $[x,\delta\lambda]$ also belong to σ_o for all δ in the interval $[1-\varepsilon,1+\varepsilon]$ for some sufficiently small positive ε. We therefore conclude that $[x',\lambda'] \varepsilon\, \sigma_1 \implies [x',\delta\lambda'] \varepsilon\, \sigma_1$ for all $\delta \,\varepsilon\, [1-\varepsilon',1+\varepsilon']$ where ε' is some sufficiently small positive real number.

We started from σ_o with co-dimension n and reached σ_1 having co-dimension n+1. The natural projection of σ_o into M was a single point; $\pi(\sigma_o) = x_o$. In other words, at σ_o the dimensional freedom along M was zero while there was n-dimensional freedom in the cotangent space. At σ_1 the dimensional freedom along M would have increased in general, i.e., $\pi(\sigma_1)$ could possibly have gathered a non-zero dimensional extension. This increase would be at the cost of dimensional freedom in the cotangent space. However, as evident from the discussion in the preceding paragraph, there always exists a one dimensional freedom, at least in the cotangent space.

The optimal path from every point in σ_1 is along the integral curve of the vector field X_2. Let the optimal path from the point $[x_1,\lambda_1] \varepsilon\, \sigma_1$ meet the boundary ∂C_2 again in time $t_2 > 0$. If no finite t_2 exists then we conclude as before that the optimal control, as we traverse the optimal

path from $[x_o, \lambda_o]$, suffers just one discontinuous switching, the switching being at the point $[x_1, \lambda_1]$. This again is trivial, we therefore assume the existence of a finite $t_2 > 0$. Let

$$[x_2, \lambda_2] = \Phi_2(t_2, [x_1, \lambda_1]) = \Phi_2(t_2, \Phi_1(t_1, [x_o, \lambda_o]))$$

$$\varepsilon\ \partial C_2 .$$

The optimal path crosses over from the control region C_2 into the control region C_3 (say) at $[x_2, \lambda_2]$. Our assumptions imply that the optimal control switches discontinuously here and $[x_2, \lambda_2]\ \varepsilon\ \mathrm{Int}^M(\bar{C}_{23})$. We again stipulate that

$$\Phi_2(t, \sigma_1) \pitchfork \mathrm{Int}^M(\bar{C}_{23})\ \forall\ t > 0. \tag{3.3}$$

Consider the intersection

$$\sigma_2 = \Phi_2(t_2, \sigma_1) \cap \mathrm{Int}^M(\bar{C}_{23}),$$

σ_2 is non-empty as $[x_2, \lambda_2]\ \varepsilon\ \sigma_2$ and it follows from (3.3) that it is a submanifold of T^*M with

$$\mathrm{codim}\ \sigma_2 = n+2 .$$

We will continue to have similar situations every time the optimal path crosses over from one control region to another if at every stage we stipulate the diffeomorphic image of σ_1 to be transversal to the corresponding open subset of the boundary. For, suppose the above situation holds at the $(m-1)$th stage $(m \leq n-1)$

– i.e., traversing the optimal path we have reached the point $[x_{m-1}, \lambda_{m-1}]$ in the manner,

$$[x_{m-1}, \lambda_{m-1}] = \Phi_{m-1}(t_{m-1}, \Phi_{m-2}(t_{m-2}, \ldots, \Phi_1(t_1, [x_o, \lambda_o]) \ldots)),$$

$$t_1 > 0,\ t_2 > 0, \ldots, t_{m-1} > 0,$$

$[x_{m-1}, \lambda_{m-1}] \in \sigma_{m-1}$,

$\sigma_{m-1} \subseteq \mathrm{Int}^M(\bar{C}_{m-1\ m})$ is a submanifold of T^*M with

codim $\sigma_{m-1} = n+(m-1)$;

and at $[x_{m-1}, \lambda_{m-1}]$, where the optimal path crosses over from C_{m-1} to C_m, the optimal control suffers the (m-1)th discontinuous switching

– we now show that similar situation holds in the mth stage.

The optimal path from every point in σ_{m-1} is along the integral curve of the vector field X_m. Let the optimal path from the point $[x_{m-1}, \lambda_{m-1}]$ meet the boundary ∂C_m again in time $t_m > 0$. If no finite t_m exists, we conclude as before that the optimal control, as we traverse the optimal path from $[x_o, \lambda_o]$, suffers (m-1) switchings only. This again is trivial; we therefore assume the existence of a finite $t_m > 0$. Let

$$\begin{aligned}[x_m, \lambda_m] &= \Phi_m(t_m, [x_{m-1}, \lambda_{m-1}]) \\ &= \Phi_m(t_m, \Phi_{m-1}(t_{m-1}, \ldots, \Phi_1(t_1, [x_o, \lambda_o]) \ldots)) \\ &\in \partial C_m .\end{aligned}$$

The optimal path crosses over from the control region C_m into the control region C_{m+1} (say) at $[x_m, \lambda_m]$. Our assumptions imply that the optimal control switches discontinuously here and $[x_m, \lambda_m] \in \mathrm{Int}^M(\bar{C}_{m\ m+1})$. We stipulate :

The System (1.1) is such that

$$\Phi_m(t, \sigma_{m-1}) \pitchfork \mathrm{Int}^M(\bar{C}_{m\ m+1}) \quad \forall \quad t > 0. \qquad (3.4)$$

If now we write

$$\sigma_m = \Phi_m(t_m, \sigma_{m-1}) \cap \mathrm{Int}^M(\bar{C}_{m\ m+1})$$

then σ_m is non-empty as $[x_m, \lambda_m] \in \sigma_m$ and from (3.4) it follow that it is a submanifold of T^*M with

$$\mathrm{codim}\ \sigma_m = n+m.$$

This process, however, cannot continue indefinitely. Consider the (n-1)th stage, we have reached the point $[x_{n-1}, \lambda_{n-1}]$ and $\sigma_{n-1} \subseteq \mathrm{Int}^M(\bar{C}_{n-1\ n})$ has codimension 2n-1. The optimal path from $[x_{n-1}, \lambda_{n-1}]$ is along the integral curve of the vector field X_n and under our stipulation either of the following and none else is true :

(i) the optimal path from $[x_{n-1}, \lambda_{n-1}]$ does not meet ∂C_n again,

(ii) the optimal path from $[x_{n-1}, \lambda_{n-1}]$ meets ∂C_n again at $[x_n, \lambda_n] \in \mathrm{Int}^M(\bar{C}_{n\ n+1})$ and that

$$\Phi_n(t, \sigma_{n-1}) \pitchfork \mathrm{Int}^M(\bar{C}_{n\ n+1}) \quad \forall \quad t > 0.$$

Now suppose (ii) is true, then σ_n has zero dimensions. This, however, is not possible. For, as we have seen before, $[x,\lambda] \varepsilon \sigma_m \Longrightarrow [x,\delta\lambda] \varepsilon \sigma_m$ for all $\delta \varepsilon [1-\varepsilon, 1+\varepsilon]$ and for all possible m, where ε is a sufficiently small positive number. Hence (i) is true, in other words the optimal path from the point $[x_{n-1}, \lambda_{n-1}]$ onwards remains in the interior of the control region C_n. We therefore conclude that the optimal control, as we traverse the optimal path from $[x_o, \lambda_o]$, does not suffer more than (n-1) discontinuous switchings.

At the outset of this analysis we assumed that $[x_o, \lambda_o]$ belongs to the interior of the control region C_1. Now suppose $[x_o, \lambda_o]$ lies on the boundary of some control region. We then traverse the optimal path from $[x_o, \lambda_o]$ for a negative time and reach a point $[x_o', \lambda_o']$, say, belonging to the interior of some control region. We denote this control region by C_1 and take $[x_o', \lambda_o']$ as our starting point. The optimal control as we traverse the optimal path from $[x_o', \lambda_o']$ does not suffer more than (n-1) discontinuous switchings $\Longrightarrow$ it does not suffer more than (n-1) discontinuous switchings as we traverse the optimal path from $[x_o, \lambda_o]$.

The condition that

$$\Phi_m(t, \sigma_{m-1}) \pitchfork \mathrm{Int}^M(\bar{C}_{m\ m+1}) \quad \forall \quad t > 0$$

for every possible m will be called the transversality condition I. We have therefore proved the following :

would be nice if one had to check for m

Consider an optimal path of the System (1.1). Suppose, every time the optimal path crosses over from a control region C_i to a control region C_j it does so at a point belonging to $\mathrm{Int}^M(\bar{C}_{ij})$ and that the optimal control suffers a discontinuous switching there. If the System (1.1) satisfies the transversality condition I then the optimal control, as we traverse this optimal path, does not suffer more than (n-1) discontinuous switchings.

==> The optimal control - considered as a function of time - governing the optimal trajectory (in M) which arises as the projection of the above optimal path does not suffer more than (n-1) discontinuities.

<u>Step 2</u>

Consider the mth stage in the above analysis. The optimal path crosses over from the control region C_m to the control region C_{m+1}. Now instead of supposing (as in step 1) that the optimal control suffers a discontinuous switching, let us suppose that it switches continuously as the optimal path crosses over from C_m to C_{m+1}. The optimal path as long as it lies in the control region C_m (C_{m+1}), is along some integral curve of the vector field X_m (X_{m+1}). Let it be possible to find a C^1 vector field $X_{m\ m+1}$ defined everywhere on $\mathrm{Int}(C_m) \cup \mathrm{Int}(C_{m+1}) \cup \mathrm{Int}^M(\bar{C}_{m\ m+1})$ such that

$$X_{m\ m+1}([x,\lambda]) = \begin{cases} X_m([x,\lambda]) & \text{when } [x,\lambda] \in C_m \\ X_{m+1}([x,\lambda]) & \text{when } [x,\lambda] \in C_{m+1} \end{cases}$$

Clearly, $X_{m\ m+1}$ can be extended smoothly up to $[x_{m-1}, \lambda_{m-1}]$. The optimal path from $[x_{m-1}, \lambda_{m-1}]$ onwards is now along an integral curve of the vector field $X_{m\ m+1}$ and it (the optimal path) continues to be so even after it has crossed over to the control region C_{m+1}. Now $[x_m, \lambda_m]$ is the point where the optimal path crosses over from C_m to C_{m+1} and by our assumption (for Step 2)

$$[x_m, \lambda_m] \in \mathrm{Int}^M(\bar{C}_{m\ m+1}).$$

It is therefore possible to carry over σ_{m-1} diffeomorphically past $C_{m\ m+1}$ into the interior of the control region C_{m+1}, where the diffeomorphism in question is generated by the same vector field - viz., $X_{m\ m+1}$ - which generates the optimal paths in the two control regions C_m and C_{m+1}. On the other hand, if it is not possible to find the C^1 vector field $X_{m\ m+1}$, it is also not possible to transport σ_{m-1} diffeomorphically into the control region C_{m+1} in the manner suggested above and we encounter a situation similar to that in step 1 where the switchings were all discontinuous. Let E^m and E^{m+1} be the extremal surfaces that represent the optima control in $\mathrm{Int}(C_m)$ and $\mathrm{Int}(C_{m+1})$.

We say that the optimal control switches non-smoothly from E^m to E^{m+1} at $[x_m, \lambda_m]$ if it is not possible to find the vector field $X_{m\ m+1}$.

Note that non-smooth switchings not only include the type of continuous switchings just discussed but also include

all discontinuous switchings. Therefore, to continue the process of reasoning as developed in step 1 it becomes logically necessary to make the following stipulation :

The System (1.1) is such that

$$\Phi_m(t,\sigma_{m-1}) \pitchfork \mathrm{Int}^M(\bar{C}_{m\ m+1}) \quad \forall \quad t > 0$$

whenever the optimal control switches non-smoothly at $[x_m,\lambda_m]$ and we write

$$\sigma_m = \Phi_m(t_m,\sigma_{m-1}) \cap \mathrm{Int}^M(\bar{C}_{m\ m+1}).$$

On the other hand if the optimal control switches smoothly at $[x_m,\lambda_m]$, i.e., if it is possible to find the vector field $X_{m\ m+1}$, we write

$$\sigma_m = \Phi_m(t_m,\sigma_{m-1}).$$

σ_m therefore has the dimensions of σ_{m-1} when the switching is smooth.

The condition that

$$\Phi_m(t,\sigma_{m-1}) \pitchfork \mathrm{Int}^M(\bar{C}_{m\ m+1}) \quad \forall \quad t > 0$$

for every possible m for which the optimal control switches non-smoothly at $[x_m,\lambda_m]$ will be called the tranversality condition II.

It is now evident that as far as tranversality condition II goes there is no point in maintaining different identities for the two control regions C_m and C_{m+1} if the optimal control switches smoothly at every point in

$\mathrm{Int}^M(\bar{C}_{m\ m+1})$ and this is true for every pair of control regions similarly related. We also recall that $\mathrm{Int}^M(\bar{C}_{m\ m+1})$ has been assumed to be a (2n-1) dimensional submanifold and this assumption has been implicitly incorporated into the constructio of the vector field $X_{m\ m+1}$ (when it is possible to construct $X_{m\ m+1}$). If $\mathrm{Int}^M(\bar{C}_{m\ m+1})$ is found to have lesser dimensionality then by our convention the optimal control switches non-smoothly at every point in $C_{m\ m+1}$. Further we assume that given two control regions C_i and C_j the submanifolds $\mathrm{Int}^M(\bar{C}_{ij})$ and $\mathrm{Int}^M(\bar{C}_{ji})$ have the same dimensionality. These facts taken together make it necessary - in the sense that it would be unnatural otherwise - to combine (taking union) all adjacent control regions, between which it is possible to switch smoothly, into a single control region. When this is done we obtain a new partition of Z. The partition sets of this partition will again be called control regions and we will use the symbols C_i to denote them. This will create no confusion, as of now, we will be dealing with this type of control regions mostly and we reserve the term "old control regions" to be used when an occasion arises to refer back to the control regions as defined earlier. It is now evident that whenever an optimal path crosses over from one control region into another the optimal control suffers a non-smooth switching. We further assume that these control regions and the sets C_{ij} possess all the properties of the old control regions and the old C_{ij}.

We repeat all that has been done until now. In other words we traverse the optimal path from the point $[x_o, \lambda_o]$ assuming that everytime the optimal path crosses over from a control region C_i to a control region C_j it does so at a point belonging to $\mathrm{Int}^M(\bar{C}_{ij})$ and assign the same meaning to the symbols $C_1, C_2, \ldots ; \sigma_1, \sigma_2, \ldots ;$ and others, as were assigned in Step 1. We however rewrite the transversality condition II as follows :

$$\Phi_m(t, \sigma_{m-1}) \not\pitchfork \mathrm{Int}^M(\bar{C}_{m\ m+1}) \quad \forall \quad t > 0$$

for every possible m.

All switchings being non-smooth switchings now, the above condition is equivalent to the earlier transversality condition II. An analysis similar to the one developed in Step 1 then proves the following :

Consider an optimal path of the System (1.1). Suppose every time the optimal path crosses over from a control region C_i to a control region C_j it does so at a point belonging to $\mathrm{Int}^M(\bar{C}_{ij})$. If the System (1.1) satisfies the transversality condition II then the optimal control, as we traverse this optimal path, does not suffer more than (n-1) non-smooth switchings.

==> The optimal control - considered as a function of time - governing the optimal trajectory (in M) which arises as the projection of the above optimal path does not suffer more than (n-1) non-smoothnesses. Here by a non-smoothness of u(t)

we either mean a discontinuity or a non-differentiability.

In the analysis above, the sets $\bar{C}_{ij}$ were all assumed to be (2n-1)-dimensional submanifolds with boundary. The analysis suffers very little change if we deviate from this assumption and take $\bar{C}_{ij}$ to be submanifolds with lesser dimensionality. Consider the mth stage again. Let $\bar{C}_{m\ m+1}$ be a 2n-k dimensional submanifold with boundary; $1 < k < 2n$. $[x_m, \lambda_m] \in \mathrm{Int}^M(\bar{C}_{m\ m+1})$, this is in keeping with the analysis in Step 1 and Step 2. We extend the transversality condition II to this case in the most natural way by stipulating that

$$\Phi_m(t, \sigma_{m-1}) \pitchfork \mathrm{Int}^M(\bar{C}_{m\ m+1}) \quad \forall \quad t > 0$$

and write

$$\sigma_m = \Phi_m(t_m, \sigma_{m-1}) \cap \mathrm{Int}^M(\bar{C}_{m\ m+1}).$$

σ_m, clearly, assumes a codimension greater than n+m. On the whole therefore the number of non-smooth switchings of the optimal control gets reduced if the dimensionality of $\bar{C}_{m\ m+1}$, at any stage m, gets reduced. Finally, if $\bar{C}_{m\ m+1}$ fails to qualify as a manifold the above analysis does not apply.

Step 3

Here we take off all restrictions that were imposed on the system. Consider the mth stage again. The optimal path crosses over from the control region C_m into the control region C_{m+1} at $[x_m, \lambda_m]$. We now allow the point $[x_m, \lambda_m]$ to lie in $(C_{m\ m+1} - \mathrm{Int}^M(\bar{C}_{m\ m+1}))$. In other words the point

$[x_m, \lambda_m]$ lies in the boundary of $C_{m\ m+1}$ (by "boundary of C_{ij}" we mean the set$(\bar{C}_{ij} - \mathrm{Int}^M(\bar{C}_{ij}))$ for the rest of the chapter). Hence there may exist other such sets such that $[x_m, \lambda_m]$ also belongs to the boundaries of these sets. Let us for a while withdraw the symbol C_{m+1} from the analysis and consider the following situation.

Let $\Delta \subsetneq \partial C_m$ denote the set of all points of the type $[x^*, \lambda^*]$ such that $[x^*, \lambda^*]$ belongs to the boundary of each of the sets $C_{mi_1}, C_{mi_2}, \ldots, C_{mi_k}; C_{j_1 m}, C_{j_2 m}, \ldots, C_{j_p m}$ and none else and let the optimal path from the point $[x_{m-1}, \lambda_{m-1}]$ meet ∂C_m again at the point $[x_m, \lambda_m] \in \Delta$. Since Δ is common to the boundaries of the above (k+p) sets we expect to find points in Δ where the optimal path crosses over from C_m to C_{i_r}, $r = 1, 2, \ldots, k$; similarly there can also exist points in Δ at which the optimal path crosses over into C_m from C_{j_r}, $r = 1, 2, \ldots, p$. Let us denote by ΔC_{ij} the set of all points in Δ where the optimal path crosses over from C_i into C_j. In keeping with our original assumption regarding the sets C_{ij} (see page 68) we assume that $\overline{\Delta C_{ij}}$ is a smooth submanifold with boundary. Δ can now be expresse as the disjoint union

$$\Delta = \Delta C_{mi_1} \cup \Delta C_{mi_2} \cup \cdots \cup \Delta C_{mi_k} \cup \Delta C_{j_1 m} \cup \Delta C_{j_2 m} \cup \cdots \cup \Delta C_{j_p m}.$$

We carry on the analysis in the following two steps.

(1) Consider the case where $[x_m, \lambda_m] \in \mathrm{Int}^M(\overline{\Delta C}_{mi_{q_1}})$ where i_{q_1} is an element of the collection $\{i_1, i_2, \ldots, i_k\}$. We take the points in $\Delta C_{mi_{q_1}}$ to be of a different nature from the points in $\mathrm{Int}^M(\bar{C}_{mi_{q_1}})$. This in turn prompts us to consider $\Delta C_{mi_{q_1}}$ to be a separate entity in itself. To make the analysis compatible with the process of reasoning developed in Step 1 and Step 2 it then becomes necessary to stipulate that

$$\Phi_m(t, \sigma_{m-1}) \not\pitchfork \mathrm{Int}^M(\overline{\Delta C}_{mi_{q_1}}) \quad \forall \quad t > 0.$$

This condition will be called the boundary transversality condition. We write

$$\sigma_m = \Phi_m(t_m, \sigma_{m-1}) \cap \mathrm{Int}^M(\overline{\Delta C}_{mi_{q_1}})$$

and to avoid unnecessary confusion denote the control region $C_{i_{q_1}}$ by the symbol C_{m+1} instead. Clearly

$$\mathrm{codim}\ \sigma_m > (n+m)$$

for, $\mathrm{codim}\ \mathrm{Int}^M(\Delta\bar{C}_{mi_{q_1}}) > 1$.

(2) Next consider the case where $[x_m, \lambda_m] \in (\Delta C_{mi_{q_1}} - \mathrm{Int}^M(\overline{\Delta C}_{mi_{q_1}}))$. In other words the point $[x_m, \lambda_m]$ lies in the boundary of the set $\Delta C_{mi_{q_1}}$ (i.e. the set $(\overline{\Delta C}_{mi_{q_1}} - \mathrm{Int}^M(\overline{\Delta C}_{mi_{q_1}})))$. Hence there can exist other such sets,

say, $\Delta C_{mi_{q_2}}$, $\Delta C_{mi_{q_3}}, \ldots, \Delta C_{mi_{q_r}}$; $\Delta C_{j_{q_1}m}$, $\Delta C_{j_{q_2}m}, \ldots, \Delta C_{j_{q_s}m}$ — where $i_{q_1}, i_{q_2}, \ldots, i_{q_r}$ are elements in the collection $\{i_1, i_2, \ldots, i_k\}$ and $j_{q_1}, j_{q_2}, \ldots, j_{q_s}$ are elements in the collection $\{j_1, j_2, \ldots, j_p\}$ — such that $[x_m, \lambda_m]$ belongs to the boundary of each of these sets and, say, these include all such sets. Let Δ' denote the set of all points in Δ which are common to the boundaries of the above $(r+s)$ sets and none else. Then Δ' when considered with reference to the above $(r+s)$ sets is equivalently situated to Δ when considered with respect to the $(k+p)$ sets $C_{mi_1}, \ldots, C_{mi_k}$; $C_{j_1m}, \ldots, C_{j_pm}$. We therefore analyse the points in Δ' in exactly the same manner we analysed the points in Δ and write Δ' as the disjoint union

$$\Delta' = \Delta' C_{mi_{q_1}} \cup \Delta' C_{mi_{q_2}} \cup \ldots \cup \Delta' C_{mi_{q_r}}$$

$$\cup \Delta' C_{j_{q_1}m} \cup \Delta' C_{j_{q_2}m} \cup \ldots \cup \Delta' C_{j_{q_s}m} .$$

Now if $[x_m, \lambda_m] \in \mathrm{Int}^M(\overline{\Delta' C_{mi_{q_1}}})$ we rewrite boundary transversality condition as follows

$$\Phi_m(t, \sigma_{m-1}) \pitchfork \mathrm{Int}^M(\overline{\Delta' C_{mi_{q_1}}}) .$$

Note that we have considered $\Delta' C_{mi_{q_1}}$ to be a separate entity in itself. We further write

$$\sigma_m = \Phi_m(t_m,\sigma_{m-1}) \cap \mathrm{Int}^M(\overline{\Delta' C_{mi_{q_1}}})$$

and denote $C_{mi_{q_1}}$ by C_{m+1} instead. It is also clear that

$$\operatorname{codim} \sigma_m > n+m$$

for, $\operatorname{codim} \mathrm{Int}^M(\overline{\Delta' C_{mi_{q_1}}}) > 1$.

If on the other hand $[x_m,\lambda_m] \in (\Delta' C_{mi_{q_1}} - \mathrm{Int}^M(\overline{\Delta' C_{mi_{q_1}}}))$ we construct a set Δ'' from Δ' in the same way we constructed Δ' starting from Δ and so on. In all cases, however, the inequality

$$\operatorname{codim} \sigma_m > n+m$$

remains valid.

At this stage we would like to mention a somewhat different situation. Let $\Delta \subseteq \partial C_m$ be the set of all points such that $[x,\lambda] \in \Delta$ implies that $[x,\lambda]$ belongs to the boundary of each of the sets C_{mk_1}, $C_{k_1 m}$; C_{mk_2}, $C_{k_2 m}$; ... ...; C_{mk_p}, $C_{k_p m}$ and none else. For such a set Δ it is possible to find points inside it such that an optimal path when it reaches this point turns back into the control region from which it came. Let $\delta \subseteq \Delta$ be the set of all points in Δ at which the optimal path turns back into C_m. We assume that $\bar{\delta}$ is a smooth manifold with boundary. Now suppose the optimal path from $[x_{m-1},\lambda_{m-1}]$ meets ∂C_m again at the point $[x_m,\lambda_m] \in \mathrm{Int}^M(\bar{\delta})$. To be consistent with the process

of reasoning developed so far we stipulate that

$$\Phi_m(t,\sigma_{m-1}) \pitchfork \mathrm{Int}^M(\bar{\delta}) \qquad \forall \;\; t > 0 .$$

This again is the boundary transversality condition and we write

$$\sigma_m = \Phi_m(t_m,\sigma_{m-1}) \cap \mathrm{Int}^M(\bar{\delta})$$

in which case

$$\mathrm{codim}\ \sigma_m > n+m.$$

Note that we have taken δ to be a separate entity. This follows from the same logical necessity which did prompt us to consider $\Delta C_{mi_{q_1}}$ and $\Delta' C_{mi_{q_1}}$ as separate entities. Further in this case where the optimal path turns back into C_m, $C_{m+1} \equiv C_m$.

We say that the System (1.1) satisfies the transversality condition if

(i) for every possible m for which $[x_m,\lambda_m] \varepsilon \mathrm{Int}^M(\bar{C}_{m\ m+1})$ the condition

$$\Phi_m(t,\sigma_{m-1}) \pitchfork \mathrm{Int}^M(\bar{C}_{m\ m+1}) \qquad \forall \;\; t > 0$$

holds;

or if

(ii) the boundary transversality condition is satisfied.

We have finally proved the following :

If the System (1.1) satisfies the transversality condition, the optimal control, as we traverse an optimal path

the set of points optimally reachable from x possesses a non-empty interior in M-the upper bound on the number of discontinuities is a number not less than (n-1). In systems for which there exists an uncountable number of vector fields to manoeuvre with, we enjoy more freedom in trying to reach different points in M from some given point. For such systems it may be possible to have the value of the upper bound less than (n-1). We, however, in our analysis, have nowhere made provision which ensures an uncountable number of vector fields to manoeuvre with, nor does the fact that the system satisfies the transversality condition ensures it - in fact the linear system $\dot{x} = Ax + bu$, $-1 \leq u \leq 1$, for which there exist just two vector fields to manoeuvre with (see page 20) satisfies the transversality condition, under some constraints; we see this in the next section. The upper bound therefore, has to be a number greater than or equal to (n-1).

3-3 Fel'dbaum's Switching Theorem

To illustrate the conclusions obtained in Section 3-2, here we re-establish the Fel'dbaum's switching theorem [1, p.120; 2,p.143].

Consider the linear time optimal problem where the state equations are given by

$$\dot{x} = Ax + bu; \ -1 \leq u \leq 1. \qquad \text{(L)}$$

The variable $x \in \mathbb{R}^n$, A is an nxn real matrix and b is an nx1

real vector. We assume that the System (L) is controllable, i.e.,

$$\text{rank } (b, Ab, \ldots, A^{n-1}b) = n$$

Fel'dbaum's switching theorem for the linear time optimal problem for (L) asserts : when the eigenvalues of A are all real the optimal control governing any optimal traject suffers at most (n-1) discontinuities. We therefore prove the following proposition.

Proposition

Let A have all real eigenvalues. Then (L) necessarily satisfies the transversality condition as a consequence of whi the optimal control governing any optimal trajectory does not suffer more than (n-1) discontinuities.

The following two lemmas are essential to prove the proposition.

Lemma 1

Consider the homogeneous linear system

$$\dot{x} = Bx, \qquad (*)$$

where $x \in \mathbb{R}^n$ and B is an nxn real matrix. Let $\Phi_*(t,\cdot)$ denote the one parameter group of diffeomorphisms generated by the vector field Bx on $\mathbb{R}^n$. We have the following :

(i) If $S \subset \mathbb{R}^n$ is a hyperplane given by

$$\langle a, x \rangle = 0$$

where a is an nx1 vector, the diffeomorphic image $\Phi_*(t$

is also a hyperplane given by

$$\langle e^{-B^T t} a, x \rangle = 0.$$

(ii) If S_1 and S_2 are two hyperplanes in $\mathbb{R}^n$ with $S_1 \pitchfork S_2$ then

$$\Phi_*(t,S_1) \pitchfork \Phi_*(t,S_2) \quad \forall\ t.$$

Proof : (i) Consider the adjoint equation

$$\dot{y} = -B^T y \tag{**}$$

and let $x(t)$ and $y(t)$ satisfy (*) and (**) respectively. Then the scalar product

$$\langle x(t), y(t)\rangle = \text{constant}.$$

For, $\frac{d}{dt} \langle x(t),y(t)\rangle = \langle \dot{x}(t),y(t)\rangle + \langle x(t),\dot{y}(t)\rangle$

$= \langle Bx(t),\ y(t)\rangle - \langle x(t), B^T\ y(t)\rangle = 0.$ Solving Equation (**) with $y(0) = a$ we obtain

$$y(t) = e^{-B^T t} a.$$

Now as $\langle a,x\rangle = 0$ for all $x \in S$, it follows that

$$\langle e^{-B^T t} a, x\rangle = 0$$

for all $x \in \Phi_*(t,S)$.

(ii) Since $\Phi_*(t,S_1)$ and $\Phi_*(t,S_2)$ are hyperplanes they must either coincide or be transversal. However, if they do coincide, the fact that (*) has a unique solution through any initial point will be violated and hence $\Phi_*(t,S_1)$ and $\Phi_*(t,S_2)$ are necessarily transversal.

Lemma 2.

Let A have all real eigenvalues. For any choice of $m \leq n$ distinct times $t_1 < t_2 < \dots < t_m$ and m integers $m_j > 0$ with $\sum_{j=1}^{m} m_j = n$, the n vectors $e^{At_j} A^{k-1} b$ $(1 \leq k \leq m_j, 1 \leq j \leq m)$ are linearly independent.

See [18] for a proof.

Proof of the Proposition :

For the System (L) the state space $M \equiv \mathbb{R}^n$, $T^*M \equiv \mathbb{R}^{2n}$ and for $x \in M$, $M_x^* = \mathbb{R}^n$. The co-state variable λ satisfies the equation

$$\dot{\lambda} = -A^T \lambda . \qquad (L')$$

The Hamiltonian is given by

$$H = \langle\lambda, Ax\rangle + \langle\lambda, b\rangle u.$$

Z, the set of all points in T^*M through which there exists an optimal path, is given by

$$Z = \{[x,\lambda] \in \mathbb{R}^{2n} : \lambda \neq 0 \text{ and } \langle\lambda, Ax\rangle + |\langle\lambda, b\rangle| \geq 0\}.$$

The optimal control is bang-bang and takes values +1 and -1 alternately. Z is therefore partitioned into two control regions C_1 and C_2 where the optimal control remains constant at the values +1 and -1 respectively. In the relative topolo of Z, C_1 and C_2 have a common boundary Δ_1 given by

$$\Delta_1 = \{[x,\lambda] \in Z : \langle b,\lambda\rangle = 0\}.$$

Further

$$\text{Int}(C_1) = \{[x,\lambda] \in Z : \langle b,\lambda\rangle > 0\}$$

$$\text{Int}(C_2) = \{[x,\lambda] \in Z : \langle b,\lambda\rangle < 0\}.$$

We take a point $[x_o,\lambda_o] \in \text{Int}(C_1)$ and study the number of switchings that the optimal control suffers as we traverse the optimal path through $[x_o,\lambda_o]$.

The portion of an optimal path lying in C_1 is nothing but an integral curve of the system of equations

$$\frac{dx}{dt} = Ax + b \qquad (L_+)$$

$$\frac{d\lambda}{dt} = -A^T\lambda \qquad (L')$$

Let X_+ be the vector field defined on T^*M by the rule

$$X_+([x,\lambda]) = \begin{bmatrix} Ax + b \\ -A^T\lambda \end{bmatrix}$$

and let $\Phi_+(t,\cdot)$ be the one parameter group of diffeomorphisms generated by X_+. Similarly the portion of an optimal path lying in C_2 is nothing but an integral curve of the system of equations

$$\frac{dx}{dt} = Ax - b \qquad (L_-)$$

$$\frac{d\lambda}{dt} = -A^T\lambda \qquad (L')$$

Let X_- be the corresponding vector field defined on T^*M and $\Phi_-(t,\cdot)$ be the one parameter group of diffeomorphisms generat by X_-.

The optimal path from $[x_o,\lambda_o]$ first crosses over (if at all it crosses over) from C_1 to C_2 then from C_2 to C_1 and so on.

Let $[x',\lambda'] \varepsilon \Delta_1$. To see which way the optimal path crosses at $[x',\lambda']$ we have to examine the Lie derivative $L_{X_+} H_-$ or equivalently $L_{X_-} H_+$ at this point, where

$$H_+ = <\lambda,Ax> + <\lambda,b> \text{ and}$$

$$H_- = <\lambda,Ax> - <\lambda,b> .$$

Now

$$L_{X_+} H_- = <\frac{\partial H_-}{\partial x} , \frac{dx}{dt}> + <\frac{\partial H_-}{\partial \lambda} , \frac{d\lambda}{dt}> .$$

Here $\frac{dx}{dt}$ and $\frac{d\lambda}{dt}$ are along the integral curve of X_+. We therefore have

$$\frac{dx}{dt} = Ax + b \text{ and } \frac{d\lambda}{dt} = -A^T \lambda .$$

Hence

$$\begin{aligned} L_{X_+} H_- &= <A^T\lambda,Ax+b> + <Ax-b,-A^T\lambda> \\ &= 2 <\lambda,Ab> \\ &= -L_{X_-} H_+ \end{aligned}$$

Similarly

$$L_{-X_+} H_- = -2 <\lambda,Ab> = -L_{-X_-} H_+$$

where

$$-X_+([x,\lambda]) = \begin{bmatrix} -Ax - b \\ A^T \lambda \end{bmatrix}$$

and $$-X_{-}([x,\lambda]) = \begin{bmatrix} -Ax + b \\ A^T\lambda \end{bmatrix}$$

From the condition of controllability of (L) it follows that the two vectors

$$\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0(nth) \\ b \end{bmatrix} \varepsilon\, \mathbb{R}^{2n} \quad \text{and} \quad \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0(nth) \\ Ab \end{bmatrix} \varepsilon\, \mathbb{R}^{2n}$$

are linearly independent. The hyperplanes

$$\langle\lambda,b\rangle = 0 \text{ and } \langle\lambda,Ab\rangle = 0$$

therefore intersect transversally. Let $[x',\lambda']$ be a point in Δ_1 such that $\langle\lambda',Ab\rangle > 0$. At this point we then have

$$L_{X_-} H_+ < 0 < L_{X_+} H_- \qquad (*)$$

and

$$L_{-X_+} H_- < 0 < L_{-X_-} H_+ \,. \qquad (**)$$

From the Relation $(*)$ it follows that H_+ decreases as we traverse the integral curve of X_- from the point $[x',\lambda']$ onwards and that H_- increases as we traverse the integral curve of X_+ from the same point, $[x',\lambda']$, onwards. Maximum principle therefore dictates that the optimal path from the point $[x',\lambda']$ onwards and close to it is along the integral curve of the vector field X_-. Similarly it follows from Relation $(**)$ that we reach the point $[x',\lambda']$ along the optimal path which is nothing but an integral curve of the

vector field X_+. In other words the optimal path crosses over from the control region C_1 into the control region C_2 at $[x',\lambda']$. We therefore have the following :

In the relative topology of Δ_1

$$Int(C_{12}) = \{[x,\lambda] \in \Delta_1 : \langle\lambda,Ab\rangle > 0\}$$

$$Int(C_{21}) = \{[x,\lambda] \in \Delta_1 : \langle\lambda,Ab\rangle < 0\}$$

and these two sets have the common boundary

$$\Delta_2 = \{[x,\lambda] \in \Delta_1 : \langle\lambda,Ab\rangle = 0\} .$$

What happens to an optimal path when it reaches a point belonging to Δ_2 can be ascertained by examining the higher order Lie derivatives as said before. These derivatives, as can be easily verified, are given by

$$L^m_{X_+} H_- = (-1)^{m+1} 2\langle\lambda,A^m b\rangle = -L^m_{X_-} H_+$$

and

$$L^m_{-X_+} H_- = -2 \langle\lambda,A^m b\rangle = -L^m_{-X_-} H_+$$

where

$$L^m_X = L_X(L^{m-1}_X).$$

The hyperplanes

$$\langle\lambda,b\rangle = 0, \ \langle\lambda,Ab\rangle = 0 \text{ and } \langle\lambda,A^2b\rangle = 0$$

intersect transversally (thanks to the controllability of (L

We therefore write the following

$$F_1 = \{[x,\lambda] \in \Delta_2 : \langle\lambda,A^2b\rangle > 0\}$$

$$G_1 = \{[x,\lambda] \in \Delta_2 : \langle\lambda, A^2b\rangle < 0\}$$

and

$$\Delta_3 = \{[x,\lambda] \in \Delta_2 : \langle\lambda, A^2b\rangle = 0\}.$$

Clearly Δ_2 is the disjoint union

$$\Delta_2 = F_1 \cup G_1 \cup \Delta_3.$$

Let $[x_a,\lambda_a] \in F_1$. Then at $[x_a,\lambda_a]$ we have

$$L_{X_+} H_- = L_{X_-} H_+ = L_{-X_+} H_- = L_{-X_-} H_+ = 0$$

$$L^2_{X_+} H_- < 0 < L^2_{X_-} H_+$$

$$L^2_{-X_+} H_- < 0 < L^2_{-X_-} H_+$$

A little consideration now shows that we reach the point $[x_a,\lambda_a]$ along the optimal path which is nothing but an integral curve of the vector field X_+ and that from the point $[x_a,\lambda_a]$ onwards the optimal path continues to be along the integral curve of X_+. In other words at $[x_a,\lambda_a]$ the optimal path turns back into the same region, viz. $\mathrm{Int}(C_1)$ from which it came and this is true for all points in F_1. Similarly at all points in G_1 the optimal path turns back into the region $\mathrm{Int}(C_2)$ from which it came. To see whi[ch] way the optimal path turns when it reaches a point belonging to Δ_3 we have to partition Δ_3 in the same way as we partitioned Δ_2. If we continue this process we obtain a partition of Δ_2 in the sense that Δ_2 can be expressed as the disjoint union

$$\Delta_2 = F_1 \cup G_1 \cup F_2 \cup G_2 \cup \cdots \cup F_{n-2} \cup G_{n-2}$$

where

$$\begin{aligned} F_k = \{[x,\lambda] \in \Delta_2 : \langle\lambda, A^2 b\rangle &= \langle\lambda, A^3 b\rangle = \cdots \\ &= \langle\lambda, A^k b\rangle = 0 \\ &\text{and } \langle\lambda, A^{k+1} b\rangle > 0\} \end{aligned}$$

and

$$\begin{aligned} G_k = \{[x,\lambda] \in \Delta_2 : \langle\lambda, A^2 b\rangle &= \langle\lambda, A^3 b\rangle = \cdots \\ &= \langle\lambda, A^k b\rangle = 0 \\ &\text{and } \langle\lambda, A^{k+1} b\rangle < 0\} \end{aligned}$$

At a point $[x,\lambda] \in F_k$ we have

$$\begin{aligned} 0 &= L_{X_+} H_- = L^2_{X_+} H_- = \cdots = L^k_{X_+} H_- \\ &= L_{X_-} H_+ = L^2_{X_-} H_+ = \cdots = L^k_{X_-} H_+ \\ &= L_{-X_+} H_- = L^2_{-X_+} H_- = \cdots = L^k_{-X_+} H_- \\ &= L_{-X_-} H_+ = L^2_{-X_-} H_+ = \cdots = L^k_{-X_-} H_+ \end{aligned}$$

and

$$L^{k+1}_{X_+} H_- < 0 < L^{k+1}_{X_-} H_+$$

if k is odd,

$$L^{k+1}_{X_-} H_+ < 0 < L^{k+1}_{X_+} H_-$$

if k is even

and finally

$$L_{-X_+}^{k+1} H_- < 0 < L_{-X_-}^{k+1} H_+ .$$

Hence if k is odd the optimal path turns back into the region $Int(C_1)$ from which it came at all points $[x,\lambda] \varepsilon F_k$. On the other hand the optimal path crosses over from $Int(C_1)$ into $Int(C_2)$ at all $[x,\lambda] \varepsilon F_k$ if k is even. Similarly at points $[x,\lambda] \varepsilon G_k$ the optimal path turns back into $Int(C_2)$ from which it came if k is odd whereas for even k it crosses over from $Int(C_2)$ into $Int(C_1)$ at points $[x,\lambda] \varepsilon G_k$.

Now the switchings encountered along an optimal path of the System (L) are all discontinuous switchings. The fact that (L) satisfies the transversality condition will therefore be established in the following two steps.

Step 1. First we suppose that every time the optimal path from $[x_o,\lambda_o]$ crosses over from C_1 to C_2 or from C_2 to C_1, it does so at points belonging to $Int(C_{12})$ or $Int(C_{21})$ (interior in the relative topology of Δ_1) respectively and show that (L) satisfies the transversality condition I.

Step 2. Next we allow the optimal path to pass through points belonging to Δ_2 and show that (L) satisfies the boundary transversality condition.

Before proceeding along the lines indicated above we note that the state and co-state equations are decoupled everywhere,

i.e., the Equations (L') are decoupled from the Equations (L_+) and also from the Equations (L_-). Let $\Sigma = C_1 \cap M^*_{x_o}$ and let σ_o be a neighbourhood of $[x_o,\lambda_o]$ in the relative topology of $M^*_{x_o}$ such that $\sigma_o \subset \Sigma$. Consider the set $\Phi_+(t_1,\sigma_o)$, let p be any point in this set then $\pi(p) = x_1$ where $x_1 \in M$ is that point which is reached in time t_1, from the point x_o, traversing the trajectory of the system of Equations (L_+) defined on M. Hence

$$\Phi_+(t_1,\sigma_o) \subsetneq M^*_{x_1}.$$

We can think of the set $\Phi_+(t_1,\sigma_o)$ in the following alternative way. Consider the co-state space $M^*_{x_o}$. Let X' be a vector field defined on $M^*_{x_o}$ by the rule $X'(\lambda) = -A^T\lambda$ and let $\Phi'(t,\cdot)$ denote the corresponding one parameter group of diffeomorphisms. Now $\sigma_o \subsetneq M^*_{x_o}$, construct the set

$$\Phi'(t_1,\sigma_o) \subsetneq M^*_{x_o}.$$

$\Phi_+(t_1,\sigma_o)$ considered as a subset of $M^*_{x_1}$ is equivalent in all respects to $\Phi'(t_1,\sigma_o)$ considered as a subset of $M^*_{x_o}$. Let Δ_1' be the subset of $M^*_{x_o}$ defined by

$$\Delta_1' = \{\lambda \in M^*_{x_o} : \langle\lambda,b\rangle = 0\}.$$

Then

$$\Phi_+(t_1,\sigma_o) \pitchfork \Delta_1$$

if $\quad \Phi'(t_1,\sigma_o) \pitchfork \Delta_1'$ (transversal as subsets of $M^*_{x_o}$)

Let us write

$$\sigma_1 = \Phi_+(t_1,\sigma_o) \cap \Delta_1$$

$$\sigma_1' = \Phi'(t_1,\sigma_o) \cap \Delta_1'$$

and, say, σ_1 and σ_1' are nonempty. σ_1 considered as a subset of $M^*_{x_1}$ is equivalent in all respects to σ_1' considered as a subset of $M^*_{x_o}$. Construct the set $\Phi_-(t_2,\sigma_1)$. For every point $p \in \Phi_-(t_2,\sigma_1)$ we have $\pi(p) = x_2$ where x_2 is that point in M which is reached in time t_2, from the point x_1, traversing the trajectory of the system of equations (L_-) defined on M. We therefore have

$$\Phi_-(t_2,\sigma_1) \subseteq M^*_{x_2}.$$

Again as before $\Phi_-(t_2,\sigma_1)$ considered as a subset of $M^*_{x_2}$ is equivalent in all respects to $\Phi'(t_2,\sigma_1')$ considered as a subset of $M^*_{x_o}$, furthermore

$$\Phi_-(t_2,\sigma_1) \pitchfork \Delta_1$$

if $\quad \Phi'(t_2,\sigma_1') \pitchfork \Delta_1'$ (transversal as subsets of $M^*_{x_o}$)

and so on.

We therefore consider the following equivalent problem.

In the n-dimensional space $\mathbb{R}^n$ of the co-state vector λ consider the following subsets

$$\Delta_1' = \{\lambda \in \mathbb{R}^n \ : \ \langle\lambda, b\rangle = 0\}$$

$$\mathrm{Int}(C_1') = \{\lambda \in \mathbb{R}^n \ : \ \langle\lambda, b\rangle > 0\}$$

$$\mathrm{Int}(C_2') = \{\lambda \in \mathbb{R}^n \ : \ \langle\lambda, b\rangle < 0\}.$$

We have, no doubt, created some confusion by defining $\mathrm{Int}(C_1')$ and $\mathrm{Int}(C_2')$ without defining the sets C_1' and C_2' a priori. The two sets as defined are however open subsets of $\mathbb{R}^n$, we could have denoted them by any other symbol, but have chosen the symbols $\mathrm{Int}(C_1')$ and $\mathrm{Int}(C_2')$ to emphasize the fact that the manner in which they are related to the equivalent problem is similar to the manner in which the sets $\mathrm{Int}(C_1)$ and $\mathrm{Int}(C_2)$ are related to the original problem. $\mathrm{Int}(C_1')$ and $\mathrm{Int}(C_2')$ have as their common boundary the set Δ_1'.

We also write

$$\Delta_2' = \{\lambda \in \Delta_1' \ : \ \langle\lambda, Ab\rangle = 0\}$$

$$\mathrm{Int}(C_{12}') = \{\lambda \in \Delta_1' \ : \ \langle\lambda, Ab\rangle > 0\}$$

$$\mathrm{Int}(C_{21}') = \{\lambda \in \Delta_1' \ : \ \langle\lambda, Ab\rangle < 0\}.$$

$\mathrm{Int}(C_{12}')$ and $\mathrm{Int}(C_{21}')$ are $(n-1)$-dimensional submanifolds in $\mathbb{R}^n$ and have Δ_2' as their common topological boundary in the relative topology of Δ_1'.

Let X' be the vector field defined on $\mathbb{R}^n$ by the rule

$$X'(\lambda) = -A^T\lambda$$

and let $\Phi'(t,\cdot)$ be the one parameter group of diffeomorphisms generated by X'.

If $\lambda \varepsilon \operatorname{Int}(C'_{12})$ then

$$<-A^T\lambda, b> = -<\lambda, Ab> < 0.$$

This implies that the integral curve of X' through λ crosses over from $\operatorname{Int}(C'_1)$ to $\operatorname{Int}(C'_2)$ at λ. Similarly the integral curve of X' crosses over from $\operatorname{Int}(C'_2)$ into $\operatorname{Int}(C'_1)$ at points belonging to $\operatorname{Int}(C'_{21})$. At points $\lambda \varepsilon \Delta'_2$ we have $<-A^T\lambda, b> = 0$. To see which way the integral curve of X' turns at these points we first observe that along this integral curve

$$\frac{d^m\lambda}{dt^m} = (-1)^m (A^T)^m \lambda \quad \text{and}$$

$$\frac{d^m\lambda}{d(-t)^m} = (A^T)^m \lambda .$$

$$\Longrightarrow <\frac{d^m\lambda}{dt^m}, b> = (-1)^m <\lambda, A^m b> \quad \text{and}$$

$$<\frac{d^m\lambda}{d(-t)^m}, b> = <\lambda, A^m b> .$$

Next, keeping in mind the way Δ_2 was partioned, we write Δ'_2 as the disjoint union

$$\Delta'_2 = F'_1 \cup G'_1 \cup F'_2 \cup G'_2 \cup \dots \cup F'_{n-2} \cup G'_{n-2} \cup \{0\}$$

where

$$F'_k = \{\lambda \varepsilon \Delta'_2 : <\lambda, A^2 b> = <\lambda, A^3 b> = \dots$$

$$= <\lambda, A^k b> = 0$$

$$\text{and } <\lambda, A^{k+1} b> > 0\}$$

and

$$G'_k = \{\lambda \in \Delta'_2 : \langle\lambda, A^2b\rangle = \langle\lambda, A^3b\rangle = \cdots = \langle\lambda, A^kb\rangle = 0 \text{ and } \langle\lambda, A^{k+1}b\rangle < 0\}$$

Now consider a point $\lambda \in F'_k$. Then at λ we have

$$\langle\frac{d^i\lambda}{dt^i}, b\rangle = \langle\frac{d^i\lambda}{d(-t)^i}, b\rangle = 0 \text{ for } i = 1,2,\ldots,k$$

$$\langle\frac{d^{k+1}\lambda}{dt^{k+1}}, b\rangle \begin{cases} > 0 \text{ if } k \text{ is odd} \\ < 0 \text{ if } k \text{ is even} \end{cases}$$

and

$$\langle\frac{d^{k+1}\lambda}{d(-t)^{k+1}}, b\rangle > 0.$$

It therefore follows that at points in F_k the integral curve of X' turns back into $\mathrm{Int}(C'_1)$ from which it came if k is odd and crosses over into $\mathrm{Int}(C'_2)$ from $\mathrm{Int}(C'_1)$ if k is even. Similarly at points in G_k the integral curve of X' turns back into $\mathrm{Int}(C'_2)$ from which it came if k is odd and crosses over from $\mathrm{Int}(C'_2)$ into $\mathrm{Int}(C'_2)$ if k is even.

Step 1. Recall that the point $[x_o, \lambda_o] \in \mathrm{Int}(C_1)$, this implies $\lambda_o \in \mathrm{Int}(C'_1)$. We first assume that the integral curve of X' through λ_o never meets Δ'_2, in other words we establish transversality condition I as a first step towards establishing transversality condition. Let σ'_o be a neighbourhood of λ_o such that $\sigma'_o \subseteq \mathrm{Int}(C'_1)$ and let it be small enough such that

its size does not hinder the arguments that follow.

We start from λ_o and traverse the integral curve of X'. Let this integral curve meet Δ_1' for the first time in time $t_1 > 0$. If no finite t_1 exists then the transversality condition is trivially satisfied and the optimal control as we traverse the optimal path (of the System (L)) from $[x_o, \lambda_o]$ onwards suffers no switchings. This immediately implies that along the optimal trajectory, from $x_o \in M$, which arises as the projection of this optimal path, the optimal control suffers no discontinuity. This is trivial. We therefore assume the existence of a finite $t_1 > 0$. Let

$$\lambda_1 = \Phi'(t_1, \lambda_o) \in \Delta_1'$$

$$\lambda_1 \neq 0 \quad \text{since} \quad \lambda_o \neq 0.$$

Evidently $\lambda_1 \in \mathrm{Int}(C_{12}')$ and the integral curve of X' crosses over from $\mathrm{Int}(C_1')$ into $\mathrm{Int}(C_2')$ at the point λ_1. Since σ_o' is an n-dimensional submanifold of $\mathbb{R}^n$, we have

$$\Phi'(t, \sigma_o') \pitchfork \mathrm{Int}(C_{12}') \quad \forall\, t > 0 \tag{3.5}$$

Consider the intersection

$$\sigma_1' = \Phi'(t_1, \sigma_o') \cap \mathrm{Int}(C_{12}').$$

σ_1' is non-empty as $\lambda_1 \in \sigma_1'$ and it follows from (3.5) that it is a submanifold of $\mathbb{R}^n$ with

$$\operatorname{codim} \sigma_1' = 1 .$$

Further σ_1' is an open subset of Δ_1'. (For two subsets A and B in $\mathbb{R}^n$, we say that A is an open subset of B if $A \subsetneq B$ and is open in the relative topology of B).

Let the integral curve of X' starting from λ_1 meet Δ_1' again in time $t_2 > 0$. Non-existence of a positive finite t_2 is again trivial. In other words starting from λ_0 and traversing the integral curve of X' we meet Δ_1' for the second time at λ_2 in time t_1+t_2.

$$0 \neq \lambda_2 = \Phi'(t_2,\lambda_1) = \Phi'(t_1+t_2,\lambda_0) \varepsilon \Delta_1'$$

Evidently $\lambda_2 \varepsilon \mathrm{Int}(C_{21}')$ and the integral curve of X' crosses over from $\mathrm{Int}(C_2')$ into $\mathrm{Int}(C_1')$ at λ_2. Now for $t > 0$, $\Phi'(t,\Delta_1')$ is the hyperplane given by the equation

$$<e^{At}b,\lambda> = 0.$$

Since the vectors $e^{At}b$ and b are linearly independent (Lemma 2) we have

$$\Phi'(t,\Delta_1') \pitchfork \Delta_1' \quad \forall \quad t > 0.$$

Since σ_1' is an open subset of Δ_1' we have

$$\Phi'(t,\sigma_1') \pitchfork \Delta_1' \quad \forall \quad t > 0.$$

Again since $\mathrm{Int}(C_{21}')$ is an open subset of Δ_1' we have

$$\Phi'(t,\sigma_1') \pitchfork \mathrm{Int}(C_{21}') \quad \forall \quad t > 0. \tag{3.6}$$

Consider the intersection

$$\sigma_2' = \Phi'(t_2,\sigma_1') \cap \mathrm{Int}(C_{21}').$$

σ_2' is non-empty as $\lambda_2 \in \sigma_2'$ and it follows from (3.6) that it is a submanifold of $\mathbb{R}^n$ with

$$\operatorname{codim} \sigma_2' = 2.$$

Further σ_2' is an open subset of

$$\Phi'(t_2,\Delta_1') \cap \Delta_1'.$$

We will continue to have similar situations every time the integral curve of X' through λ_o meets Δ_1'. For, suppose the above situation holds at the (m-1)th stage, we take m to be odd and less than or equal to n-1

– i.e., the integral curve of X' from λ_o meets Δ_1' for the (m-1)th time at λ_{m-1} in the manner

$$0 \neq \lambda_{m-1} = \Phi'(t_1+t_2+\dots+t_{m-1},\lambda_o) \in \Delta_1' ,$$
$$t_1 > 0,\ t_2 > 0,\dots,t_{m-1} > 0;$$

$$\lambda_{m-1} \in \operatorname{Int}(C_{21}'),$$

(recall that m is odd); the integral curve of X' crosses over from $\operatorname{Int}(C_2')$ into $\operatorname{Int}(C_1')$ at λ_{m-1};

$$\Phi'(t,\sigma_{m-2}') \pitchfork \operatorname{Int}(C_{21}') \quad \forall \quad t > 0,$$

$$\sigma_{m-1}' = \Phi'(t_{m-1}, \sigma_{m-2}') \cap \operatorname{Int}(C_{21}'),$$

$$\lambda_{m-1} \in \sigma_{m-1}' ,$$

σ_{m-1}' is a submanifold of $\mathbb{R}^n$ with

$$\operatorname{codim} \sigma_{m-1}' = m-1$$

and finally σ'_{m-1} is an open subset of

$$\Phi'(t_2+t_3+\dots+t_{m-1}, \Delta_1') \cap \Phi'(t_3+t_4+\dots+t_{m-1}, \Delta_1')$$
$$\cap \dots \cap \Phi'(t_{m-1}, \Delta_1') \cap \Delta_1'$$

- we show that similar situation holds in the mth stage.

Let the integral curve of X' starting from λ_{m-1} meet Δ_1' again in time $t_m > 0$ at the point λ_m. If no finite t_m exists then the transversality condition I is satisfied and the optimal control as we traverse the optimal path of the System (L) from the point $[x_o, \lambda_o]$ suffers (m-1) discontinuous switchings only. This immediately implies that along the optimal trajectory through x_o, which arises as the projection of this optimal path, the optimal control suffers (m-1) discontinuities only. This is trivial, we therefore, assume the existence of a finite $t_m > 0$. We have

$$0 \neq \lambda_m = \Phi'(t_1+t_2+\dots+t_m, \lambda_o) \in \Delta_1'$$

Evidently $\lambda_m \in \mathrm{Int}(C'_{12})$ and the integral curve of X' crosses over from $\mathrm{Int}(C_1')$ into $\mathrm{Int}(C_2')$ at λ_m. Now, for $t > 0$ the m hyperplanes

$$\Phi'(t_2+t_3+\dots+t_{m-1}+t, \Delta_1'),$$
$$\Phi'(t_3+t_4+\dots+t_{m-1}+t, \Delta_1'),$$
$$\dots\dots\dots\dots,$$
$$\dots\dots\dots\dots,$$
$$\dots\dots\dots\dots,$$
$$\Phi'(t, \Delta_1'),$$
$$\Delta_1'$$

are represented by the following m equations - in the same order -

$$\langle e^{A(t_2+t_3+\cdots+t_{m-1}+t)} b, \lambda\rangle = 0,$$
$$\langle e^{A(t_3+t_4+\cdots+t_{m-1}+t)} b, \lambda\rangle = 0,$$
$$\cdots\cdots\cdots\cdots,$$
$$\cdots\cdots\cdots\cdots,$$
$$\cdots\cdots\cdots\cdots,$$
$$\langle e^{At} b, \lambda\rangle = 0,$$
$$\langle b, \lambda\rangle = 0.$$

Now from lemma 2 it follows that the m vectors

$$e^{A(t_2+t_3+\cdots+t_{m-1}+t)} b,$$
$$e^{A(t_3+t_4+\cdots+t_{m-1}+t)} b,$$
$$\cdots\cdots\cdots\cdots,$$
$$\cdots\cdots\cdots\cdots,$$
$$\cdots\cdots\cdots\cdots,$$
$$e^{At} b,$$
$$b$$

are linearly independent. We therefore have

$$[\Phi'(t_2+t_3+\cdots+t_{m-1}+t, \Delta_1') \cap \Phi'(t_3+t_4+\cdots+t_{m-1}+t, \Delta_1')$$
$$\cap \cdots \cap \Phi'(t, \Delta_1')] \not\subset \Delta_1' \quad \forall\ t > 0.$$

However, $\Phi'(t,\sigma_{m-1}')$ is an open subset of

$$\Phi'(t_2+t_3+\cdots+t_{m-1}+t, \Delta_1') \cap \Phi'(t_3+t_4+\cdots+t_{m-1}+t, \Delta_1')$$
$$\cap \cdots \cap \Phi'(t,\Delta_1').$$

Hence

$$\Phi'(t,\sigma_{m-1}') \not\subset \Delta_1' \quad \forall\ t > 0.$$

Again since $\mathrm{Int}(C'_{12})$ is an open subset of Δ'_1 we have

$$\Phi'(t,\sigma'_{m-1}) \pitchfork \mathrm{Int}(C'_{12}) \quad \forall\, t > 0 . \qquad (3.7)$$

Consider the intersection

$$\sigma'_m = \Phi'(t_m, \sigma'_{m-1}) \cap \mathrm{Int}(C'_{12}) .$$

σ'_m is non-empty as $\lambda_m \varepsilon \sigma'_m$ and it follows from (3.7) that it is a submanifold of $\mathbb{R}^n$ with

$$\mathrm{codim}\ \sigma'_m = m.$$

Further σ'_m is an open subset of

$$\Phi'(t_2+t_3+\dots+t_m, \Delta'_1) \cap \Phi'(t_3+t_4+\dots+t_m, \Delta'_1)$$
$$\cap \dots \cap \Phi'(t_m, \Delta'_1) \cap \Delta'_1 .$$

Note : If m is taken to be an even integer the analysis remains essentially the same, except for the fact that at λ_m the integral curve of X' crosses over from $\mathrm{Int}(C'_2)$ into $\mathrm{Int}(C'_1)$.

We now show that the integral curve of the vector field X' starting from λ_o cannot meet Δ'_1 more than (n-1) times. For, suppose it meets Δ'_1 for the nth time at λ_n in the manner

$$0 \neq \lambda_n = \Phi'(t_1+t_2+t_3+\dots+t_n, \lambda_o) \ \varepsilon\ \Delta'_1 ,$$
$$t_1 > 0,\ t_2 > 0, \dots,\ t_n > 0.$$

From the preceding analysis it is clear that λ_n belongs to the intersection

$$\Phi'(t_2+t_3+\dots+t_n\ ,\ \Delta_1') \cap \Phi'(t_3+t_4+\dots+t_n\ ,\ \Delta_1')$$
$$\cap \dots \cap \Phi'(t_n\ ,\Delta_1') \cap \Delta_1' .$$

However, as the n vectors

$$e^{A(t_2+t_3+\dots+t_n)}\, b,\ e^{A(t_3+t_4+\dots+t_n)}\, b,$$
$$\dots,\ e^{At_n}\, b,\ b \qquad \text{are linearly independent,}$$

the above intersection contains the single vector 0. But $\lambda_n \neq 0$, we therefore reach a contradiction. Hence the integral curve of X' through λ_o meets Δ_1' for at most (n-1) times. Now, recall that the starting point λ_o belongs to Int(C_1'). If $[x_o,\lambda_o]$ were taken to be a point inside Int(C_2), λ_o would belong to Int(C_2') instead. This, however induces little change into the analysis just concluded. On the other hand, if $[x_o,\lambda_o]$ were taken to be a point on Δ_1, λ_o would belong to Δ_1'. In such cases, as indicated before, we traverse the optimal path from $[x_o,\lambda_o]$ for a negative time until we reach a point belonging to the interior of some control region and take it as our starting point. This is equivalent to traversing the integral curve of X' through $\lambda_o \varepsilon \Delta_1'$ for a negative duration until we reach a point, say, λ_o' belonging to either Int(C_1') or Int(C_2') and take this λ_o' as our starting point. The integral curve of X' from λ_o' onwards does not meet Δ_1' more than (n-1) times ==> the integral curve of X' from λ_o onwards does not meet Δ_1' more than (n-2) times. These facts in turn imply the following :

Consider an optimal path of the System (L). Suppose every time this optimal path crosses over from C_1 to C_2 or from C_2 to C_1 it does so at points belonging to $\mathrm{Int}(C_{12})$ or $\mathrm{Int}(C_{21})$ respectively, then the optimal control as we traverse this optimal path does not suffer more than (n-1) discontinuous switchings.

==> The optimal control - considered as a function of time - governing the optimal trajectory (in M) which arises as the projection of this optimal path does not suffer more than (n-1) discontinuities.

Step 2.

We now take off the restriction that was imposed on the integral curve of X' through λ_0 viz., that it never meets Δ_2'. Consider the mth stage of Step 1 again. Suppose the integral curve of X' from the point λ_{m-1} meets Δ_1' again in time $t_m > 0$ at a point $\lambda_m \in \Delta_2' \subset \Delta_1'$. We have to show that the boundary transversality condition is satisfied. Let us go back to the set Δ_2 for a while. We have partitioned this set by expressing it as a disjoint union

$$\Delta_2 = F_1 \cup G_1 \cup F_2 \cup G_2 \cup \cdots \cup F_{n-2} \cup G_{n-2}.$$

Let k be an integer with $1 \leq k \leq n-2$. In keeping with the convention as adopted in Step 3 of Section (3-2) we take the points in F_k to be of a different nature from the points in $\mathrm{Int}(C_{12})$, $\mathrm{Int}(C_{21})$; $F_1, F_2, \ldots, F_{k-1}, F_{k+1}, \ldots, F_{n-2}; G_1, G_2, \ldots, G_{n-2}$

and a similar convention holds for the points in G_k. This in turn prompts us to treat each of the sets $F_1, F_2, \ldots, F_{n-2}$; $G_1, G_2, \ldots, G_{n-2}$ as a separate entity. To take care of this fact we have partitioned Δ_2' in a similar manner by expressing it as the disjoint union

$$\Delta_2' = F_1' \cup G_1' \cup F_2' \cup G_2' \cup \ldots \cup F_{n-2}' \cup G_{n-2}' \cup \{0\}.$$

Now suppose $\lambda_m \in F_k \subset \Delta_2'; 1 \leq k \leq n-2$. (If $\lambda_m \in G_k$ instead, the corresponding analysis is similar to the one that follows). The boundary transversality condition is satisfied if

$$\Phi'(t, \sigma_{m-1}') \pitchfork F_k' \quad \forall\, t > 0.$$

Let Q_i represent the hyperplane

$$\langle\lambda, A^i b\rangle = 0; \qquad i = 0,1,2,\ldots$$

F_k is evidently an open subset of

$$Q_0 \cap Q_1 \cap \ldots \cap Q_k$$

and from the analysis in Step 1 it follows that $\Phi'(t,\sigma_{m-1}')$ is an open subset of

$$\Phi'(t_2+t_3+\ldots+t_{m-1}+t,\Delta_1') \cap \Phi'(t_3+t_4+\ldots$$

$$\ldots+ t_{m-1}+t,\Delta_1') \cap \ldots \cap \Phi'(t,\Delta_1').$$

Further from Lemma 2 it follows that the (m+k) vectors

$$e^{A(t_2+t_3+\cdots+t_{m-1}+t)}\, b,$$
$$e^{A(t_3+t_4+\cdots+t_{m-1}+t)}\, b,$$
$$\vdots \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \vdots\,,$$
$$e^{At}\, b,$$
$$b\,,$$
$$Ab\,,$$
$$\cdots,$$
$$\cdots,$$
$$A^k\, b,$$

are linearly independent. Hence

$$\Phi'(t,\sigma'_{m-1}) \pitchfork F'_k \qquad \forall\, t > 0\ .$$

We write

$$\sigma'_m = \Phi'(t_m\,,\sigma'_{m-1}) \cap F'_k\ ,$$

$\lambda_m \in \sigma'_m$, codim $\sigma'_m = m+k$ and σ'_m is an open subset of

$$\Phi'(t_2+t_3+\cdots+t_m\,,\Delta'_1) \cap \Phi'(t_3+t_4+\cdots$$
$$\cdots+t_m\,,\Delta'_1) \cap \cdots \cap \Phi'(t_m\,,\Delta'_1)$$
$$\cap\, Q_o \cap Q_1 \cap \cdots \cap Q_k\ . \tag{3.8}$$

Note that the intersection (3.8) contains the single vector 0 if $m \geq n-k$. Since λ_m cannot be 0 we conclude that the integral curve of X' from λ_{m-1} can never meet Δ'_1 at a point belonging to $F'_k \subset \Delta'_1$ if $m \geq n-k$ or for that matter it can

never meet Δ_1' at points belonging to any of the sets $F_{k+1},\ldots,F_{n-2}$; $G_k, G_{k+1},\ldots,G_{n-2}$ when $m \geq n-k$.

The proof of the transversality condition can now be completed by invoking Lemma 2 everytime the integral curve of X' from λ_m onwards meets Δ_1'. Further since codim $\sigma_m = m+k$ an argument similar to the one developed in Step 1 shows that the integral curve of X' from λ_m onwards does not meet Δ_1' for more than $(n-1)-(m+k) = n-m-(k+1)$ times and that the number of times it meets Δ_1' gets further reduced if in its course of journey it passes through Δ_2' once again or more number of times. In other words starting from λ_o this integral curve meets Δ_1' at most $n-(k+1)$ times.

We have therefore proved that given any $\lambda_o \varepsilon \mathbb{R}^n$ the integral curve of X' from λ_o does not meet Δ_1' more than $(n-1)$ times.

==> The optimal control as we traverse an optimal path of the System (L) does not suffer more than (n-1) discontinuous switchings.

==> The optimal control - considered as a function of time - governing any optimal trajectory of the System (L) does not suffer more than (n-1) discontinuities.

The Feldbaum's switching theorem is therefore established and with this we conclude our study of the discontinuities of the optimal control.

APPENDIX -A

Let E be an extremal surface. A reviewer has pointed out that the projection $\Psi: E \longrightarrow \Psi(E)$ can be many-to- one in spite of the arguments given in pp.37-38. Let it be so. Let $A \subseteq \Psi(E)$ be such that E represents the optimal control at every point in A. If $A = \emptyset$ then E is of no use to us. Consider the subset $E' \subseteq E$ defined by

$$E' = \{([x, \lambda], u) \in E : [x, \lambda] \in A \text{ and } u \text{ gives the value of the optimal control at } [x, \lambda]\}.$$

In other words E' is that portion of E which actually represents the optimal control in A. Now, since corresponding to each point $[x, \lambda] \in Z$ there exists a unique value of the optimal control it follows that the projection $\Psi: E' \longrightarrow A$ is one-to-one. Therefore even though the projection of E into T^*M is many-to-one, the projection of the portion which is responsible for representing the optimal control is one-to-one. Our sole interest in an extremal surface lies in the fact that it represents the optimal control some where. We can therefore assume that $\Psi: E \longrightarrow \Psi(E)$ is one-to-one without loss of generality.

BIBLIOGRAPHY

1. L.S. Pontryagin, V.G. Boltyanskii, R.V. Gamkrelidze, and E.F. Mishchenko, The Mathematical Theory of Optimal Processes, Interscience Publishers, Inc., New York, 1962.

2. E.B. Lee, L. Markus, Foundations of Optimal Control Theory, John Wiley and Sons, Inc., New York, 1967.

3. M. Spivak, A Comprehensive Introduction to Differential Geometry, Vol. 1, Publish or Perish, Inc., Berkeley, California, 1970.

4. V. Guillemin, A. Pollack, Differential Topology, Prentice Hall, Inc., New Jersey, 1974.

5. M. Golubitsky, V. Guillemin, Stable Mappings and Their Singularities, Springer-Verlag, New York, Inc., 1973.

6. M.M. Vainberg, and V.A. Trenogin, Theory of Branching of Solutions of Non-Linear Equations, Noordhoff International Publishing, Leyden, 1974.

7. D.F. Lawden, Optimal Trajectories for Space Navigation, Butterworth, London, 1963.

8. C. Martin, R. Hermann, Editors, The 1976 Ames Research Centre (NASA) Conference on Geometric Control Theory, Math. Sci. Press, Brookline, Mass., 1977.

9. C. Marchal, Transfers Optimaux Entre Orbites Elliptiques (Durée Indifférente), Ph.D. Thesis, University of Paris, June 1967.

10. H.W. Small, Minimum Fuel Time-Free Transfer Between Elliptic Orbits, Ph.D. Dissertation, Stanford University, Aug. 1972.

11. R.W. Brockett, System Theory on Group Manifolds and Coset Spaces, SIAM J. Control, 10(1972), pp. 265-284.

12. R.W. Brockett, Nonlinear Systems and Differential Geometry, Proc. IEEE, 64 (1976), pp. 61-72.

13. H.J. Sussmann and V. Jurdjevic, Controllability of Nonlinear Systems, J. Differential Equations, 12 (1972), pp. 95-116.

14. A.J. Krener, A Generalization of Chow's Theorem and The Bang-Bang Theorem to Nonlinear Control Problems, SIAM J. Control, 12 (1974), pp. 43-52.

15. N. Levitt and H.J. Sussmann, On Controllability by means of Two Vector Fields, SIAM J. Control, 13 (1975), pp. 1271-1281.

16. H.J. Sussmann, On the Number of Directions Needed to Achieve Controllability, SIAM J. Control 13(1975), pp. 414-419.

17. H.J. Sussmann, Bounds on the Number of Switchings for Trajectories of Piecewise Analytic Vector Fields, J. Differential Equations, 43 (1982), pp. 399-418.

18. O. Hájek, Terminal Manifolds and Switching Locus, Mathematical Systems Theory, 6(1973), pp. 289-301.

19. H.J. Sussmann, A Bang-Bang Theorem with Bounds on the Number of Switchings, SIAM J. Control and Optimization, 17 (1979), pp. 629-651.

20. T.N. Edelbaum, How Many Impulses? Astronautics and Aeronautics, 5(1967), pp. 64-69.